

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:54:58 ; Search time 44 Seconds
(without alignments)

Title: US-09-471-255-2

perfect score: 1039

Sequence: 1 MKFSKKYIAAGSAVTLSL.....IELRLPSCGEVIKKNLSDFIA 1039

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 8

Total number of hits satisfying chosen parameters: 169

Minimum DB seq length: 0

Minimum DB seq length:	Maximum DB seq length:
200000000	200000000

Post-processing: Listing first 500 summaries

Database : A_Geneseq_101002:*

1: /SID2/cgdata/geneseg/geneseq-emb1/AA1980.DAT.*
2: /SID2/cgdata/geneseg/geneseq-emb1/AA1981.DAT.*
3: /SID2/cgdata/geneseg/geneseq-emb1/AA1982.DAT.*
4: /SID2/cgdata/geneseg/geneseq-emb1/AA1983.DAT.*
5: /SID2/cgdata/geneseg/geneseq-emb1/AA1984.DAT.*
6: /SID2/cgdata/geneseg/geneseq-emb1/AA1985.DAT.*
7: /SID2/cgdata/geneseg/geneseq-emb1/AA1986.DAT.*
8: /SID2/cgdata/geneseg/geneseq-emb1/AA1987.DAT.*
9: /SID2/cgdata/geneseg/geneseq-emb1/AA1988.DAT.*
10: /SID2/cgdata/geneseg/geneseq-emb1/AA1989.DAT.*
11: /SID2/cgdata/geneseg/geneseq-emb1/AA1990.DAT.*
12: /SID2/cgdata/geneseg/geneseq-emb1/AA1991.DAT.*
13: /SID2/cgdata/geneseg/geneseq-emb1/AA1992.DAT.*
14: /SID2/cgdata/geneseg/geneseq-emb1/AA1993.DAT.*
15: /SID2/cgdata/geneseg/geneseq-emb1/AA1994.DAT.*
16: /SID2/cgdata/geneseg/geneseq-emb1/AA1995.DAT.*
17: /SID2/cgdata/geneseg/geneseq-emb1/AA1996.DAT.*
18: /SID2/cgdata/geneseg/geneseq-emb1/AA1997.DAT.*
19: /SID2/cgdata/geneseg/geneseq-emb1/AA1998.DAT.*
20: /SID2/cgdata/geneseg/geneseq-emb1/AA1999.DAT.*
21: /SID2/cgdata/geneseg/geneseq-emb1/AA2000.DAT.*
22: /SID2/cgdata/geneseg/geneseq-emb1/AA2001.DAT.*
23: /SID2/cgdata/geneseg/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1039	100.0	1039	21	AA812715		Streptococcus pneu
2	1039	100.0	1039	23	AAU75932		Streptococcus pneu
3	1019	98.1	1019	21	AA812722		Streptococcus pneu
4	1019	98.1	1019	21	AA812751		Streptococcus pneu
5	1019	98.1	1019	23	AAU84021		truncated variant
6	814	78.3	1019	23	AA812750		Streptococcus pneu
7	807	77.7	807	21	AA833997		Truncated variant
8	807	77.7	1152	23	AAU84054		S. pneumoniae deri
9	807	77.7	1238	23	AAU84056		S. pneumoniae deri
10	807	77.7	1365	23	AAU84057		S. pneumoniae deri

Caps 0;

QY 61 QKEGIAEQIVIKITDQGYVTSHGDRHYHYNGKVPYDALFSEELLMKDPNYQLKADIVN 120
DB 61 QKEGIAEQIVIKITDQGYVTSHGDRHYHYNGKVPYDALFSEELLMKDPNYQLKADIVN 120
QY 121 EVKGGYIIKVDGKYVYVLKDAHADNVRTKDEINRQKQEHYKDNKYNVAVARSQRY 180
DB 121 EVKGGYIIKVDGKYVYVLKDAHADNVRTKDEINRQKQEHYKDNKYNVAVARSQRY 180
QY 181 TNDGVVFNPAIDIIEDTGNAYIVPHGGHYHYTPKSDLSASELAARAKAHLAGNMMPSOLS 240
DB 181 TNDGVVFNPAIDIIEDTGNAYIVPHGGHYHYTPKSDLSASELAARAKAHLAGNMMPSOLS 240
QY 241 YSTASDNTQSVAKGTSKPANKSENQLSLKELYDPSAQRYSESGLVDFDPAKIIIS 300
DB 241 YSTASDNTQSVAKGTSKPANKSENQLSLKELYDPSAQRYSESGLVDFDPAKIIIS 300
QY 301 TNGVAIPHGDHYHFTIPYKSLSALEKATARMVPISTGSTVSTNAKPNEVYSSLSLSN 360
DB 301 TNGVAIPHGDHYHFTIPYKSLSALEKATARMVPISTGSTVSTNAKPNEVYSSLSLSN 360
QY 361 PSLATTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFFHYPKSNOIGQPTLPNNSLA 420
DB 361 PSLATTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFFHYPKSNOIGQPTLPNNSLA 420
QY 421 TSPSPILNPGTSHKHEEDGCGFDANRILIAEDESQVMSHGDHNYHFFKDLTEEQIKA 480
DB 421 TSPSPILNPGTSHKHEEDGCGFDANRILIAEDESQVMSHGDHNYHFFKDLTEEQIKA 480
QY 481 AQKHLBEVTSNGLDLSLSHEDQDYPGNKAKEMKDLKIEERKAGIMQYGVKRESIVVN 540
DB 481 AQKHLBEVTSNGLDLSLSHEDQDYPGNKAKEMKDLKIEERKAGIMQYGVKRESIVVN 540
QY 541 KKNALIIYPHGDHHDPTDEHKPVGIGHSHSNYELFKPEEGVAKGKNVYTGELINV 600
DB 541 KKNALIIYPHGDHHDPTDEHKPVGIGHSHSNYELFKPEEGVAKGKNVYTGELINV 600
QY 601 VLLKSTNNQNTLANGOKRVSFPPLEKKGICINMLVLIIPDGKVLKGVKGVKFG 660
DB 601 VLLKSTNNQNTLANGOKRVSFPPLEKKGICINMLVLIIPDGKVLKGVKGVKFG 660
QY 661 EGVNTANFELDQYLPQGTFFKTYTASQDYPEVSDGTFTVPTSLAYKMASOTIYPPHA 720
DB 661 EGVNTANFELDQYLPQGTFFKTYTASQDYPEVSDGTFTVPTSLAYKMASOTIYPPHA 720
QY 721 GDYLVNPFQFVPGKTDALRVFDFHGNAYLENNYKVGELKLPKLNQGTTRTAGNK 780
DB 721 GDYLVNPFQFVPGKTDALRVFDFHGNAYLENNYKVGELKLPKLNQGTTRTAGNK 780
QY 781 IPVTFMANAYLDNQSTYIIVEPPILEKENQTDKPSILPQPKRKAQENSKLDEKVEPKTS 840
DB 781 IPVTFMANAYLDNQSTYIIVEPPILEKENQTDKPSILPQPKRKAQENSKLDEKVEPKTS 840
QY 841 EKVEKEKLSGTGNTSNLSTLEEVPTDVPQEVAKFAESYGMKLENVLFNMDGTIELYLP 900
DB 841 EKVEKEKLSGTGNTSNLSTLEEVPTDVPQEVAKFAESYGMKLENVLFNMDGTIELYLP 900
QY 901 SGSEVIRKKNMADFTGEAPQNGENKPSNGKSVSTGTVENOPTENKPADSLPEAPNEKPVKP 960
DB 901 SGSEVIRKKNMADFTGEAPQNGENKPSNGKSVSTGTVENOPTENKPADSLPEAPNEKPVKP 960
QY 961 ENSTDNGMLNPEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDLSVIFNMDGTI 1020
DB 961 ENSTDNGMLNPEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDLSVIFNMDGTI 1020
QY 1021 ELRLPSGEVIKKNLSDFIA 1039
DB 1021 ELRLPSGEVIKKNLSDFIA 1039

RESULT 2
AAU75932
ID AAU75932 standard; Protein; 1039 AA.

XX AAU75932;
AC 08-MAY-2002 (first entry)
XX Streptococcus pneumoniae BVH-3 protein version #1.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection.
XX Streptococcus pneumoniae.
OS WO200198334-A2.
XX 27-DEC-2001.
XX 19-JUN-2001; 2001WO-CA00908.
XX 20-JUN-2000; 2000US-212683P.
PR (SHIR-) SHIRE BIOCHEM INC.
PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI: 2002-122272/16.
XX N-PSDB: ABK15101.
DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia.
XX Example 1; Fig 6; 113pp; English.
PS The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocidia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This is the amino acid sequence of
CC Streptococcus pneumoniae protein BVH-3, used to create the antigenic
CC peptides described in the method of the invention.
XX Sequence 1039 AA;
SQ Query Match 100.0%; Score 1039; DB 23; Length 1039;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1039; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKFSKKYTAAGSAIVLSLCAYALNQHSQENKNNRVSYVDGSSQSKSENLTDPQVS 60
DB 1 MKFSKKYTAAGSAIVLSLCAYALNQHSQENKNNRVSYVDGSSQSKSENLTDPQVS 60
QY 61 QKEGIAEQIVIKITDQGYVTSHGDRHYHYNGKVPYDALFSEELLMKDPNYQLKADIVN 120
DB 61 QKEGIAEQIVIKITDQGYVTSHGDRHYHYNGKVPYDALFSEELLMKDPNYQLKADIVN 120
QY 121 EVKGGYIIKVDGKYVYVLKDAHADNVRTKDEINRQKQEHYKDNKYNVAVARSQRY 180
DB 121 EVKGGYIIKVDGKYVYVLKDAHADNVRTKDEINRQKQEHYKDNKYNVAVARSQRY 180

```
QY 181 TTDGYYFNPADIIEDTGNAYIVPHGGHYHYIPKSDLSASELAALAAKHAHLAKGNMQPSQLS 240
DB 181 TTDGYYFNPADIIEDTGNAYIVPHGGHYHYIPKSDLSASELAALAAKHAHLAKGNMQPSQLS 240
QY 241 YSSTASDNNTQSVAKGTSKPKANKSENLOSLKELYDPSAORYSESGLVDPDAKIISR 300
DB 241 YSSTASDNNTQSVAKGTSKPKANKSENLOSLKELYDPSAORYSESGLVDPDAKIISR 300
QY 301 TPNQVAIPRGDHYHFTIPYKLSALEPKIARMVPISTGTVSTNAKPNVSVSLGSLSSN 360
DB 301 TPNQVAIPRGDHYHFTIPYKLSALEPKIARMVPISTGTVSTNAKPNVSVSLGSLSSN 360
QY 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPNNLSA 420
DB 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPNNLSA 420
QY 421 TSPSPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSHGDNHNYFFPKDLTEEQIKA 480
DB 421 TSPSPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSHGDNHNYFFPKDLTEEQIKA 480
QY 481 AQKHELVKTSKSHGLDLSHSHEDYPCNAKEMKDLCKLEEKIAGIMKQYGVKRESIVVN 540
DB 481 AQKHELVKTSKSHGLDLSHSHEDYPCNAKEMKDLCKLEEKIAGIMKQYGVKRESIVVN 540
QY 541 KEKNALIIYPRGDDHHADPIDDKHPVGIGSHSNYELFKPEEGVAKKEGKVYTGELTNV 600
DB 541 KEKNALIIYPRGDDHHADPIDDKHPVGIGSHSNYELFKPEEGVAKKEGKVYTGELTNV 600
QY 601 VNLKSTNNQNTFLANGCKRVFSFPPLEKLGILNMLVLIITPDGKVLKVSQKVF 660
DB 601 VNLKSTNNQNTFLANGCKRVFSFPPLEKLGILNMLVLIITPDGKVLKVSQKVF 660
QY 661 EGVGNIANFELDQYLPFGTFRKTYIASKDYPEVSYDGTFTVPTSLAYKNASOTIFYPFA 720
DB 661 EGVGNIANFELDQYLPFGTFRKTYIASKDYPEVSYDGTFTVPTSLAYKNASOTIFYPFA 720
QY 721 GDYLRVNOFOFVPGKTDALVRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTITAGNK 780
DB 721 GDYLRVNOFOFVPGKTDALVRVDFEFGHNAVLENNYKVEIKLPIPKLNQGTITAGNK 780
QY 781 IPTVFMANAYLDNOSTYIIVEVPTLEKENTOKPSILPQFKRNKAQENSKLDEKVEPKTS 840
DB 781 IPTVFMANAYLDNOSTYIIVEVPTLEKENTOKPSILPQFKRNKAQENSKLDEKVEPKTS 840
QY 841 EKVEKELSTGNTSNTSLTEEVPTVDPVOEKVAKPAESYGMKLENVLEPMGDTIELYLP 900
DB 841 EKVEKELSTGNTSNTSLTEEVPTVDPVOEKVAKPAESYGMKLENVLEPMGDTIELYLP 900
QY 901 SGVIVKKNMAFTGEAPQNGENKPSKENGKSVSTGTVENOPTENKPADSLPEAPNEKPKVP 960
DB 901 SGVIVKKNMAFTGEAPQNGENKPSKENGKSVSTGTVENOPTENKPADSLPEAPNEKPKVP 960
QY 961 ENSTDGMLNPEGNVSGDPMPLDPALEAPVADVPQVEKLEKFTASYGLGLDSVIFNMDGTI 1020
DB 961 ENSTDGMLNPEGNVSGDPMPLDPALEAPVADVPQVEKLEKFTASYGLGLDSVIFNMDGTI 1020
QY 1021 ELRLPSEVITKKNLSDFIA 1039
DB 1021 ELRLPSEVITKKNLSDFIA 1039
```

RESULT 3

AAB12722

ID AAB12722 standard; Protein; 1019 AA.

XX

AC AAB12722;

XX

DT 21-NOV-2000 (first entry)

XX

DE Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.

XX

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

```
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO2000392399-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
WP; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia.
XX
PS Claim 18; Fig 20; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3M protein antigen.
XX
SQ Sequence 1019 AA;
```

Query Match 98.1%; Score 1019; DB 21; Length 1019;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CAYALNQHRSENKDNRRVSYDGSQSSQSENLTDPQVSQKEGIAEQIVIKITDQYV 80

DB 1 CAYALNQHRSENKDNRRVSYDGSQSSQSENLTDPQVSQKEGIAEQIVIKITDQYV 60

QY 81 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADATVNEVKGYLIKVDGKVVYVKLD 140

DB 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADATVNEVKGYLIKVDGKVVYVKLD 120

QY 141 AAHADNVRTKDEINRQKQEHVKDKNEKNSNVAVARSQGRYTTNDGYVFNPAITIEDTNA 200

DB 121 AAHADNVRTKDEINRQKQEHVKDKNEKNSNVAVARSQGRYTTNDGYVFNPAITIEDTNA 180

QY 201 YIVPHGHYHYIPKSDLSASELAALAAKHAHLAKGNMQPSQLSSTASDNNTQSVAKGTSK 260

DB 181 YIVPHGHYHYIPKSDLSASELAALAAKHAHLAKGNMQPSQLSSTASDNNTQSVAKGTSK 240

QY 251 PANKSENLOSLKELYDPSAORYSESGLVDPDAKIISRTPNGVAIPRGDHYHFTIPYSK 320

DB 241 PANKSENLOSLKELYDPSAORYSESGLVDPDAKIISRTPNGVAIPRGDHYHFTIPYSK 300

QY 321 LSALEEKIARMVPISTGTVSTNAKPNVSVSLGSLSSNPSSLTTSKELSSASDGYIFN 380

DB 301 LSALEEKIARMVPISTGTVSTNAKPNVSVSLGSLSSNPSSLTTSKELSSASDGYIFN 360

QY 381 PKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPNNSLATPSPSLPINPGTSKHEED 440

DB 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPNNSLATPSPSLPINPGTSKHEED 420

QY 441 GYGFDANKRIIAEDSGFVMSHGDNHNYFFKDLTEEQIKAAQKHELVKTSKSHGLDLS 500

DB 421 GYGFDANKRIIAEDSGFVMSHGDNHNYFFKDLTEEQIKAAQKHELVKTSKSHGLDLS 480

QY 501 HEQDYPGNAKEMKDLCKLEEKIAGIMKQYGVKRESIVVNEKKNALIIYPHGDHHDHPID 560

Db 481 HEQDYPGNAKEMKDLKKIEIKAGIMKQYGVKRESIVVNEKKNALIIYPHGDHHPADPID 540
Qy 561 EHKPVGIGHSHSNVELFKPECEVAKKEGNKVTGTEELTNVNNLLKSNFNNQNTFLANGQ 620
Db 541 EHKPVGIGHSHSNVELFKPECEVAKKEGNKVTGTEELTNVNNLLKSNFNNQNTFLANGQ 600
Qy 621 KRVSFSPPELEKKGILGINMLVKLITPDGKVLKESGKVGEGVGNIANFELDQYPLPQGT 680
Db 601 KRVSFSPPELEKKGILGINMLVKLITPDGKVLKESGKVGEGVGNIANFELDQYPLPQGT 660
Qy 681 FKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDAL 740
Db 661 FKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDAL 720
Qy 741 VRVDFEFGHGNAYLENNYKVGSEIKLPIPKLNQGTTRTACNKIPVTFMANAYLDNQSTIVE 800
Db 721 VRVDFEFGHGNAYLENNYKVGSEIKLPIPKLNQGTTRTACNKIPVTFMANAYLDNQSTIVE 780
Qy 801 VPILKEKNQTDKPSILPQFKENKAQENSKLDEKVEEPEKTSKEVEKELSETGNSSTL 860
Db 781 VPILKEKNQTDKPSILPQFKENKAQENSKLDEKVEEPEKTSKEVEKELSETGNSSTL 840
Qy 861 BEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
Db 841 BEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900
Qy 921 GENKPESENGKYSTGTIVENQPTENKPADSLPEAPNEKPVKPEENSTDNGLNPEGNVGSDDPM 980
Db 901 GENKPESENGKYSTGTIVENQPTENKPADSLPEAPNEKPVKPEENSTDNGLNPEGNVGSDDPM 960
Qy 981 LDPALAEAPVDPVQEKLEKTASYGLDSVIFNMMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 961 LDPALAEAPVDPVQEKLEKTASYGLDSVIFNMMDGTIELRLPSGEVIKKNLSDFIA 1019

RESULT 4
AAB12751
ID AAB12751 standard; Protein; 1019 AA.
XX AC AAB12751;
XX 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae strain SP64 BVH-3 protein antigen.
XX Streptococcus pneumoniae: BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX Streptococcus pneumoniae.
OS WO200039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA01218.
XX PR 23-DEC-1998; 98US-0113800.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI: 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX Disclosure: Fig 11; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein

CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;
Query Match 98.1%; Score 1019; DB 21; Length 1019;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 21 CAVALNOHRSQENKDNRRVSYVDGSSQSKSENLITPDQVSQKEGIAEQIIVKITQGGYV 80
Db 1 CAVALNOHRSQENKDNRRVSYVDGSSQSKSENLITPDQVSQKEGIAEQIIVKITQGGYV 60
Qy 81 TSHGDHYHYNGKVPYDALFSEELLMKDPNYQLKDADIVNEVKGYYIIKVDGKYYVYLDK 140
Db 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNYQLKDADIVNEVKGYYIIKVDGKYYVYLDK 120
Qy 141 AAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRYTTNDGVYFNPADIIEDTGN 200
Db 121 AAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRYTTNDGVYFNPADIIEDTGN 180
Qy 201 YIYPHGGHYHYIPKSDLSASELAARAHLAGKNMPSQLSYSTASDNNQTSVAKGSTK 260
Db 181 YIYPHGGHYHYIPKSDLSASELAARAHLAGKNMPSQLSYSTASDNNQTSVAKGSTK 240
Qy 261 PANKSENLOSLKELYDPSAQRYSQSDGLVFPDPAKIIISRTPNGVAIPHGDHYHFIYSK 320
Db 241 PANKSENLOSLKELYDPSAQRYSQSDGLVFPDPAKIIISRTPNGVAIPHGDHYHFIYSK 300
Qy 321 LSALAEKRIARVPISGTGSTVSTNAKPNEVVSLSGSSNPSLITTSKELSSASDGIYFN 380
Db 301 LSALAEKRIARVPISGTGSTVSTNAKPNEVVSLSGSSNPSLITTSKELSSASDGIYFN 360
Qy 381 PKDIVEETATAYIVRHGDHFIYIPKSNQIQGPTLPNNSLATPSPSLPINPGTSHERHEED 440
Db 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIQGPTLPNNSLATPSPSLPINPGTSHERHEED 420
Qy 441 GYGFANDRIITAEDESGFVMSGHGHNHYFFKDLTEQIKAAQKHLEBEVKTSHNGLSLSS 500
Db 421 GYGFANDRIITAEDESGFVMSGHGHNHYFFKDLTEQIKAAQKHLEBEVKTSHNGLSLSS 480
Qy 501 HEQDYPGNAKEMKDLKKIEIKAGIMKQYGVKRESIVVNEKKNALIIYPHGDHHPADPID 560
Db 481 HEQDYPGNAKEMKDLKKIEIKAGIMKQYGVKRESIVVNEKKNALIIYPHGDHHPADPID 540
Qy 561 EHKPVGIGHSHSNVELFKPEEGVAKKEGNKVTGTEELTNVNNLLKSNFNNQNTFLANGQ 620
Db 541 EHKPVGIGHSHSNVELFKPEEGVAKKEGNKVTGTEELTNVNNLLKSNFNNQNTFLANGQ 600
Qy 621 KRVSFSPPELEKKGILGINMLVKLITPDGKVLKESGKVGEGVGNIANFELDQYPLPQGT 680
Db 601 KRVSFSPPELEKKGILGINMLVKLITPDGKVLKESGKVGEGVGNIANFELDQYPLPQGT 660
Qy 681 FKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDAL 740
Db 661 FKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPGKTDAL 720
Qy 741 VRVDFEFGHGNAYLENNYKVGSEIKLPIPKLNQGTTRTACNKIPVTFMANAYLDNQSTIVE 800
Db 721 VRVDFEFGHGNAYLENNYKVGSEIKLPIPKLNQGTTRTACNKIPVTFMANAYLDNQSTIVE 780
Qy 801 VPILKEKNQTDKPSILPQFKENKAQENSKLDEKVEEPEKTSKEVEKELSETGNSSTL 860
Db 781 VPILKEKNQTDKPSILPQFKENKAQENSKLDEKVEEPEKTSKEVEKELSETGNSSTL 840
Qy 861 BEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
Db 841 BEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900

QY 921 GENKPSSENGKSTGTVENOPTENKPADSLPAPNPKPKPKPENSTNDGMLNPEGNGVSDPM 980
DB 901 GENKPSSENGKSTGTVENOPTENKPADSLPAPNPKPKPKPENSTNDGMLNPEGNGVSDPM 960
QY 981 LDPALFEAPAVDPVQEKLEKFTASVGLGSDSVIFNMDDGTIELRLPSGEVVKKNLSDFIA 1039
DB 961 LDPALFEAPAVDPVQEKLEKFTASVGLGSDSVIFNMDDGTIELRLPSGEVVKKNLSDFIA 1019

RESULT 5
AAU84021
ID AAU84021 standard; Peptide; 1019 AA.
XX AAU84021;
AC AAU84021;
XX 08-MAY-2002 (first entry)
XX Truncated variant of *S. pneumoniae* BVH-3, BVH-3M.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutin.
KW
XX Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
XX 27-DEC-2001.
XX 19-JUN-2001; 2001WO-CA00908.
XX 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
XX Example 1; Page -; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (i) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (ii) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
CC as Streptococcus agalactiae, *S. dysgalactiae*, *S. uberis*, *S. norcardia* or
CC *Staphylococcus aureus*) in an individual susceptible to the infection.
CC A polynucleotide (iii) encoding (i) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for *S. pneumoniae* infection. (iii) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating *S. pneumoniae* nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX Sequence 1019 AA;

Query Match 98.1%; Score 1019; DB 23; Length 1019;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CAYALNHRSEKNDNRVSYVDGSSOSQSENLTPDQVSKQEGIOAEQIVIKITDQGVY 80
DB 1 CAYALNHRSEKNDNRVSYVDGSSOSQSENLTPDQVSKQEGIOAEQIVIKITDQGVY 60
QY 81 TSHGCDHYHYNGVVPYDALFSEBLLMKDPNYOLKDDADIVNEVKGGYIIKVDGKYVYLKD 140
DB 61 TSHGCDHYHYNGVVPYDALFSEBLLMKDPNYOLKDDADIVNEVKGGYIIKVDGKYVYLKD 120
QY 141 AAHADNVRTKDEINRQKQEHVKDNEKNSNVAVARSQGRYTTNDGYVFPNADIETGNA 200
DB 121 AAHADNVRTKDEINRQKQEHVKDNEKNSNVAVARSQGRYTTNDGYVFPNADIETGNA 180
QY 201 YIVPHGCHYHYIPKSDLSASELAALAAKAGLAGKNMQPSQLSYSTASDNNTQSAKGSSTK 260
DB 181 YIVPHGCHYHYIPKSDLSASELAALAAKAGLAGKNMQPSQLSYSTASDNNTQSAKGSSTK 240
QY 261 PANKSENLOSLKELVDPSAQRYSESGLVFPDPAKIIISRTPNGVAIPIHGDHYHFIPIYSK 320
DB 241 PANKSENLOSLKELVDPSAQRYSESGLVFPDPAKIIISRTPNGVAIPIHGDHYHFIPIYSK 300
QY 321 LSALKEKIARMPITSGTSTVNAPNEVSVSLGSLSSNPSSLTTSKELSSASDGYIFN 380
DB 301 LSALKEKIARMPITSGTSTVNAPNEVSVSLGSLSSNPSSLTTSKELSSASDGYIFN 360
QY 381 PKDIVEETATAYIVRHGDHFHYIPKSNQIQOPTLPNNSLATPSPSLPIPGTSHKHEED 440
DB 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIQOPTLPNNSLATPSPSLPIPGTSHKHEED 420
QY 441 GYGFDANRIIAEDSEGFVMSHGDNHYFFKKDLTEBQIKAAQKHLEEVKTSNGLDLSLS 500
DB 421 GYGFDANRIIAEDSEGFVMSHGDNHYFFKKDLTEBQIKAAQKHLEEVKTSNGLDLSLS 480
QY 501 HEQDYPGNAKEMKDLKKIEKTAGIMKQXGVKRESVYNKEKNALIIYPHGDHHDADPID 560
DB 481 HEQDYPGNAKEMKDLKKIEKTAGIMKQXGVKRESVYNKEKNALIIYPHGDHHDADPID 540
QY 561 EHPVGGHSHSNYELFKPEEGVAKKEGKNVYTGEBLTNVVNLKKNSTFNNQFTLANGQ 620
DB 541 EHPVGGHSHSNYELFKPEEGVAKKEGKNVYTGEBLTNVVNLKKNSTFNNQFTLANGQ 600
QY 621 KRVSFSPPELEKKLGINMLVKLITPDGKYLEKSGVFGEGVGNIANFELDQYLPQGT 680
DB 601 KRVSFSPPELEKKLGINMLVKLITPDGKYLEKSGVFGEGVGNIANFELDQYLPQGT 660
QY 681 FKYTIAKSDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAVPKGTDAL 740
DB 661 FKYTIAKSDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAVPKGTDAL 720
QY 741 VRVDFPHGNAYLENNYKVGIEIKLPKLNQGTTRAGNKIPVTFMANAYLDNQSYIYE 800
DB 721 VRVDFPHGNAYLENNYKVGIEIKLPKLNQGTTRAGNKIPVTFMANAYLDNQSYIYE 780
QY 801 VPILEKENQTDKPSILPQPKRNKAQENSKLDEKVEPKTSEKVEKEKLETSNSTLS 860
DB 781 VPILEKENQTDKPSILPQPKRNKAQENSKLDEKVEPKTSEKVEKEKLETSNSTLS 840
QY 861 EEPVTDVPQVEKVAKEAESYGMKLENVLFNMDGTIELYLPSPGEVVKKNADFTGEAPQGN 920
DB 841 EEPVTDVPQVEKVAKEAESYGMKLENVLFNMDGTIELYLPSPGEVVKKNADFTGEAPQGN 900
QY 921 GENKPSSENGKSTGTVENOPTENKPADSLPAPNPKPKPKPENSTNDGMLNPEGNGVSDPM 980
DB 901 GENKPSSENGKSTGTVENOPTENKPADSLPAPNPKPKPKPENSTNDGMLNPEGNGVSDPM 960
QY 981 LDPALFEAPAVDPVQEKLEKFTASVGLGSDSVIFNMDDGTIELRLPSGEVVKKNLSDFIA 1039
DB 961 LDPALFEAPAVDPVQEKLEKFTASVGLGSDSVIFNMDDGTIELRLPSGEVVKKNLSDFIA 1019

RESULT 6
AAB12750

ID AAB12750 standard; Protein; 1019 AA.
 XX AAB12750;
 AC
 XX
 DT 21-NOV-2000 (first entry)
 DE
 DE Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.
 XX
 XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 XX 20-DEC-1999; 99WO-CA01218.
 PF
 XX 23-DEC-1998; 98US-0113800.
 PR
 XX (BIOC-) BIOCHEM PHARMA INC.
 PA
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 PI
 XX WPI: 2000-452397/39.
 DR
 XX
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 PS Disclosure; Fig 11; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the proteins
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-3 protein antigen, from the present invention.
 XX
 SQ Sequence 1019 AA;

Query Match 78.3%; Score 814; DB 21; Length 1019;
 Best Local Similarity 99.8%; Pred. NO. 0;
 Matches 1014; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 21 CAYALNHRQENKNNRVSYVDGSSQSKSENLTPOVSKREGIOAEQIVIXITDQYV 80
 DB 1 CAYALNHRQENKNNRVSYVDGSSQSKSENLTPOVSKREGIOAEQIVIXITDQYV 60

QY 81 TSHGCHYHYNGKVPYDALFSELLMKDPNQLKADIVNEVKGYYIIKVDGYYVYLKD 140
 DB 61 TSHGCHYHYNGKVPYDALFSELLMKDPNQLKADIVNEVKGYYIIKVDGYYVYLKD 120

QY 141 AAHADNVRTKDEINRQKQEHVKDNEKNSVAVARSQGYTTNDGYVFNPAIIIEDTGN 200
 DB 121 AAHADNVRTKDEINRQKQEHVKDNEKNSVAVARSQGYTTNDGYVFNPAIIIEDTGN 180

QY 201 YIVPHGGHYHYIPKSDLSASELAALAGKNNQPSQLSYSTASDNNTQSVAKGSTK 260
 DB 181 YIVPHGGHYHYIPKSDLSASELAALAGKNNQPSQLSYSTASDNNTQSVAKGSTK 240

QY 261 PANKSENQSLKELYDSPAQRYSSEGLVFPDPAKIIISRTPNGVAIPHGDDHYHFIPYSK 320
 DB 241 PANKSENQSLKELYDSPAQRYSSEGLVFPDPAKIIISRTPNGVAIPHGDDHYHFIPYSK 300

QY 321 LSALEEKIARMPISGTGTVSTNAKNEVYVSSLSGSSNPSSLTTSKELSSASDGYFN 380
 DB 301 LSALEEKIARMPISGTGTVSTNAKNEVYVSSLSGSSNPSSLTTSKELSSASDGYFN 360

QY 381 PKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPOTSHKHEED 440
 DB 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPOTSHKHEED 420

QY 441 GYGFDANRIIAEDSGFVMSHGDHNNHYEFKDLTEEQKAAQKHLVEEYKTSHGNDLSUSS 500
 DB 421 GYGFDANRIIAEDSGFVMSHGDHNNHYEFKDLTEEQKAAQKHLVEEYKTSHGNDLSUSS 480

QY 501 HBQDYPGNAKEMKLDKIEEKIAGIMKOYGVKRESIVVNEKNAIIPHGDDHHADPID 560
 DB 481 HBQDYPGNAKEMKLDKIEEKIAGIMKOYGVKRESIVVNEKNAIIPHGDDHHADPID 540

QY 561 EHKPVGIGHSHSNYELFPKEGVAKEKNKYVTGELTNVNNLKNSTFNQNTLANGQ 620
 DB 541 EHKPVGIGHSHSNYELFPKEGVAKEKNKYVTGELTNVNNLKNSTFNQNTLANGQ 600

QY 621 KRVSFSPPELEKLGINNLYKLITPDGKVLKESGVGEGVGNIAFELDQPYLPQOT 680
 DB 601 KRVSFSPPELEKLGINNLYKLITPDGKVLKESGVGEGVGNIAFELDQPYLPQOT 660

QY 681 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPGKTDAL 740
 DB 661 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPGKTDAL 720

QY 741 VRVDFEFGHAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTETMANAYLDNQSTIYVE 800
 DB 721 VRVDFEFGHAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTETMANAYLDNQSTIYVE 780

QY 801 VPILKEKNTDKPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTL 860
 DB 781 VPILKEKNTDKPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTL 840

QY 861 BEVPTVDVPOEKVAKFAESYGNKLENVLFNMDGTIELYLPSEGVKKKNMADFTGEAPOGN 920
 DB 841 BEVPTVDVPOEKVAKFAESYGNKLENVLFNMDGTIELYLPSEGVKKKNMADFTGEAPOGN 900

QY 921 GENKSENGKVGSTGVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPGNVGSDPM 980
 DB 901 GENKSENGKVGSTGVENOPTENKPADSLPEAPNEKPKVPENSTONGMLNPGNVGSDPM 960

QY 981 LDPALAEAPAVDPVQEKLEKFTASGLGLDLSVIFNMDGTIELRLPSGEVKKKNLSD 1036
 DB 961 LDPALAEAPAVDPVQEKLEKFTASGLGLDLSVIFNMDGTIELRLPSGEVKKKNLSD 1016

RESULT 7
 AAU83997
 ID AAU83997 standard; Peptide: 807 AA.
 XX
 AC AAU83997;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Truncated variant of S. pneumoniae BVH-11, NEW25.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 XX
 PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 XX 19-JUN-2001; 2001WO-CA00908.
 PF
 XX 20-JUN-2000; 2000US-212683P.
 PR
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 PI
 XX

DR WPI: 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Example 1; Page : 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX Sequence 807 AA;

Query Match 77.7%; Score 807; DB 23; Length 807;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NMQPSQLSYSTASDNNQTSVAGSKSPANKSENQSLKLLKELYDPSAQRYSSEGLVP 292
DB 1 NMQPSQLSYSTASDNNQTSVAGSKSPANKSENQSLKLLKELYDPSAQRYSSEGLVP 60
QY 293 DPAAIISRTNGVAIPGHDHYHPIPTSKLSALEEKIARMPISGTSTVSTNAKPNNEVYS 352
DB 61 DPAAIISRTNGVAIPGHDHYHPIPTSKLSALEEKIARMPISGTSTVSTNAKPNNEVYS 120
QY 353 SLGSLSNPSLTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGQP 412
DB 121 SLGSLSNPSLTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGQP 180
QY 413 TLPNNSLATPSLPIINPGTSHKHEEDGYGFDPANRIIAEDSGFVMSHGDHNYFFKKD 472
DB 181 TLPNNSLATPSLPIINPGTSHKHEEDGYGFDPANRIIAEDSGFVMSHGDHNYFFKKD 240
QY 473 LTESQIKAAQKHLEEVKTSINGLSDSSHEQDYPGNAKEMKMDKKIEKIAIMQYGV 532
DB 241 LTESQIKAAQKHLEEVKTSINGLSDSSHEQDYPGNAKEMKMDKKIEKIAIMQYGV 300
QY 533 KRESIVYVKNKNAIYPHGDHHDADIDEHHPVIGHSNHYELFEPGEVAKKEGKVVY 592
DB 301 KRESIVYVKNKNAIYPHGDHHDADIDEHHPVIGHSNHYELFEPGEVAKKEGKVVY 360
QY 593 TGEELTNVALLKNSTFNQNTLANGQKRVSFSPPELEKLGIMNKLIPDQGVLE 652
DB 361 TGEELTNVALLKNSTFNQNTLANGQKRVSFSPPELEKLGIMNKLIPDQGVLE 420
QY 653 KVSCKVGEVGNIANFELDQVLPQGTFTYTIASKDYPEVSDGTFVTPTSLAYKMSQ 712
DB 421 KVSCKVGEVGNIANFELDQVLPQGTFTYTIASKDYPEVSDGTFVTPTSLAYKMSQ 480
QY 713 TIFVPPFHAGDTYLRVNPQFAPVPGTDAVRFDFEFGHGNAYLNNYKVGELKLPKPLNQG 772
DB 481 TIFVPPFHAGDTYLRVNPQFAPVPGTDAVRFDFEFGHGNAYLNNYKVGELKLPKPLNQG 540

QY 773 TTRTAGNKIPVTEMANAYLDNQSTYIVVEVPILEKENOTDKPSILPQFKRNKAQENSKLDE 832
DB 541 TTRTAGNKIPVTEMANAYLDNQSTYIVVEVPILEKENOTDKPSILPQFKRNKAQENSKLDE 600
QY 833 KVEEPTSEKVEKEKLSSETGNSTNSTLEEYPTVDPVOEKVAKFAESYGMKLENVFNMD 892
DB 601 KVEEPTSEKVEKEKLSSETGNSTNSTLEEYPTVDPVOEKVAKFAESYGMKLENVFNMD 660
QY 893 GTIELYLSGGEVKKKNMADFTGEAPQNGENKPSNGKSVTGTVENQPTENKPADSLPEA 952
DB 661 GTIELYLSGGEVKKKNMADFTGEAPQNGENKPSNGKSVTGTVENQPTENKPADSLPEA 720
QY 953 PNEKPKVPKPNSTNDGMLNPEGNVSGDPMPLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1012
DB 721 PNEKPKVPKPNSTNDGMLNPEGNVSGDPMPLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 780
QY 1013 IFNMDGTIELRLPSPGEVKKKLSDFIA 1039
DB 781 IFNMDGTIELRLPSPGEVKKKLSDFIA 807
RESULT 8
ASU84054
ID AU84054 standard; Peptide: 1152 AA.
AC AU84054;
XX
DT 08-MAY-2002 (first entry)
XX
DE S. pneumoniae derived chimeric peptide, NEW27.
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
XX BVH-11-2.
OS Streptococcus pneumoniae.
QS Synthetic.
XX WO200198334-A2.
XX
PD 27-DEC-2001.
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX (SHTR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Quellet C, Charland N, Martin D, Brodeur B;
WPI: 2002-122272/16.
XX
DR
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Example 1; Page : 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful in a diagnostic
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX

SQ Sequence 1152 AA;
Query Match 77.7%; Score 807; DB 23; Length 1152;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NNQPSQLSYSTASDNTQSVAKGSTSPANKSENLOSLKELYDSPSAQRYSESGLVLF 292
DB 346 NNQPSQLSYSTASDNTQSVAKGSTSPANKSENLOSLKELYDSPSAQRYSESGLVLF 405
QY 293 DPAKIISRTTPNGVAIPHGDHYHFIYPSKLSALEEKIARMVPISTGTVSTNAKNEVVS 352
DB 406 DPAKIISRTTPNGVAIPHGDHYHFIYPSKLSALEEKIARMVPISTGTVSTNAKNEVVS 465
QY 353 SLGSSNPSSTTSKELSSASDGIYFNPKDIVEETATAYIVRHGDHFIYIPKSNQIQCP 412
DB 466 SLGSSNPSSTTSKELSSASDGIYFNPKDIVEETATAYIVRHGDHFIYIPKSNQIQCP 525
QY 413 TLPNNSLATPSPSLNPOTSHKEEDGYGPDANRIIAEDSGFVMSHGDHNVFFKKD 472
DB 526 TLPNNSLATPSPSLNPOTSHKEEDGYGPDANRIIAEDSGFVMSHGDHNVFFKKD 585
QY 473 LTEEQIKAAQKHEEVKTSNHLGDLSSHEQDYPGNAKEMKDLKIEKIAGIMQKQGV 532
DB 586 LTEEQIKAAQKHEEVKTSNHLGDLSSHEQDYPGNAKEMKDLKIEKIAGIMQKQGV 645
QY 533 KRESIVVNEKNALIIYPHGDHHAADIDHKKPVGTCHSHSNVLEKPEGVAKKEGKVY 592
DB 646 KRESIVVNEKNALIIYPHGDHHAADIDHKKPVGTCHSHSNVLEKPEGVAKKEGKVY 705
QY 593 TGEELTNVNLKNSFNQNTFLANGQKRVSFSPPELEKKGILGIMLVKLTIPDGKYLE 652
DB 706 TGEELTNVNLKNSFNQNTFLANGQKRVSFSPPELEKKGILGIMLVKLTIPDGKYLE 765
QY 653 KVSQKVGEGVGNIANFELDQYLPQGTFTYIAASKDYPEVSDGTFTVPTSLAYKMASQ 712
DB 766 KVSQKVGEGVGNIANFELDQYLPQGTFTYIAASKDYPEVSDGTFTVPTSLAYKMASQ 825
QY 713 TIFYPHAGDTVLRVNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGSEIKLPIPKLNQ 772
DB 826 TIFYPHAGDTVLRVNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGSEIKLPIPKLNQ 885
QY 773 TTRTAGNKIPVTFMANAYLDNSTIYVEVPILEKENQTDKPSILQFQRNKAQENSKLDE 832
DB 886 TTRTAGNKIPVTFMANAYLDNSTIYVEVPILEKENQTDKPSILQFQRNKAQENSKLDE 945
QY 833 KYEPTSEKVEKEKLSSETNGSTNSTLEVTVPDQVEKVAKFESYGMKLENVLFNMD 892
DB 946 KYEPTSEKVEKEKLSSETNGSTNSTLEVTVPDQVEKVAKFESYGMKLENVLFNMD 1005
QY 893 GTIELYLPSPGEVIKKNMADFTGAPOGNGENKPSNGKYSTGTVENQPTENKPADSLPEA 952
DB 1006 GTIELYLPSPGEVIKKNMADFTGAPOGNGENKPSNGKYSTGTVENQPTENKPADSLPEA 1065
QY 953 PNEKPKPENSTNDGMLNPEGNVGSDPMLDPALEAPADVPQOEKLEKFTASYGLGLDSV 1012
DB 1066 PNEKPKPENSTNDGMLNPEGNVGSDPMLDPALEAPADVPQOEKLEKFTASYGLGLDSV 1125
QY 1013 IFNMDGTIELRPSGGEVIKKNLSDFTA 1039
DB 1126 IFNMDGTIELRPSGGEVIKKNLSDFTA 1152

RESULT 9
AAU84056
ID AAU84056 standard; Peptide; 1238 AA.
XX
AC AAU84056;
XX
DT 08-MAY-2002 (first entry)
XX
DE S. pneumoniae derived chimeric peptide, NEW29.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutain;
KW BVH-11-2.
OS Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
epitope-bearing polypeptides, useful as vaccine components for treating
or preventing streptococcal infections such as otitis media,
meningitis, and bacteraemia .
Example 1; Page : 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90%
identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
comprising (I) is useful for therapeutic or prophylactic treatment of
meningitis, otitis media, bacteraemia or pneumonia infection in an
individual susceptible to these disorders. (II) is also useful for
therapeutic or prophylactic treatment of any streptococcal bacterial
infection (e.g., caused by Streptococcus pneumoniae, group A
Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
Staphylococcus aureus) in an individual susceptible to the infection.
A polynucleotide (III) encoding (I) is useful in DNA immunisation
techniques. The Streptococcus polypeptides are useful in a diagnostic
test for S. pneumoniae infection. (III) is useful for designing DNA
probes for use in detecting the presence of Streptococcus in a biological
sample suspected of containing the bacteria. The DNA probes may also be
used for detecting circulating S. pneumonia nucleic acid in a sample for
diagnosing streptococcal infections. This sequence represents a chimeric
gene created from fragments and variant fragments of Streptococcus
pneumoniae genes, described in the method of the invention.
Note: This sequence does not appear in the specification but has
been created according to information given in the invention.

SQ Sequence 1238 AA;
Query Match 77.7%; Score 807; DB 23; Length 1238;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NNQPSQLSYSTASDNTQSVAKGSTSPANKSENLOSLKELYDSPSAQRYSESGLVLF 292
DB 432 NNQPSQLSYSTASDNTQSVAKGSTSPANKSENLOSLKELYDSPSAQRYSESGLVLF 491
QY 293 DPAKIISRTTPNGVAIPHGDHYHFIYPSKLSALEEKIARMVPISTGTVSTNAKNEVVS 352

```
Db 492 DPAKIIISRTNGVAIPHGDHYHFIYPSKLSALEEKIARMVPISGTGSTVSTNAKPNEVVS 551
Qy 353 SLGSLSNPSSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGDHFIYIPKSNQIGQP 412
Db 552 SLGSLSNPSSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGDHFIYIPKSNQIGQP 611
Qy 413 TLPNNSLATPSPLPINPGTISHEKHEEDGYGFDANRIIAEDSGFVMSHGDNHNYFFKDD 472
Db 612 TLPNNSLATPSPLPINPGTISHEKHEEDGYGFDANRIIAEDSGFVMSHGDNHNYFFKDD 671
Qy 473 LTEEQIKAAQKHLEEVKTSINGLSDLSHSDQYPGNAKEMKDLKKIEKTAGIMKQYGV 532
Db 672 LTEEQIKAAQKHLEEVKTSINGLSDLSHSDQYPGNAKEMKDLKKIEKTAGIMKQYGV 731
Qy 533 KRESIVVVKENAIYIPHGDDHHDADIDEHKPVGIGHSNHYELFKPEEGVAKKEGNKYV 592
Db 732 KRESIVVVKENAIYIPHGDDHHDADIDEHKPVGIGHSNHYELFKPEEGVAKKEGNKYV 791
Qy 593 TGEELTNVNVLLKNSTFNNQFTLANGQKRVSFSPFPELEKLGINKLVKLITPDGKYLE 652
Db 792 TGEELTNVNVLLKNSTFNNQFTLANGQKRVSFSPFPELEKLGINKLVKLITPDGKYLE 851
Qy 653 KVSCKVGEVGVGNIANFELDQVLPQGTFKYTIASKDYPEVSDGTFVTPTSLAYKMASQ 712
Db 852 KVSCKVGEVGVGNIANFELDQVLPQGTFKYTIASKDYPEVSDGTFVTPTSLAYKMASQ 911
Qy 713 TIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQ 772
Db 912 TIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQ 971
Qy 773 TTRTAGNKIPVTMANAYLDNQSTYIIVEPILKEKNOTDKPSTLPQFKRKAQENSKLDE 832
Db 972 TTRTAGNKIPVTMANAYLDNQSTYIIVEPILKEKNOTDKPSTLPQFKRKAQENSKLDE 1031
Qy 833 KVEEPTSEKVEKEKLETSQNSTLEEVPTVDPQVEKVAFAESYGMKLENVLFNMD 892
Db 1032 KVEEPTSEKVEKEKLETSQNSTLEEVPTVDPQVEKVAFAESYGMKLENVLFNMD 1091
Qy 893 GTIELYLPSSGEVTKKNADFTGEAPOGNGENKPSGKNGKSTGTVENOPTENKPADSLPEA 952
Db 1092 GTIELYLPSSGEVTKKNADFTGEAPOGNGENKPSGKNGKSTGTVENOPTENKPADSLPEA 1151
Qy 953 PNEKPVKPNSTNGMLNPGNVSDFMPLDPALEEAAPVDPVOEKLEKFTASYGLGLDSV 1012
Db 1152 PNEKPVKPNSTNGMLNPGNVSDFMPLDPALEEAAPVDPVOEKLEKFTASYGLGLDSV 1211
Qy 1013 IFNMDGTIELRLPSGEVTKKNLSDFIA 1039
Db 1212 IFNMDGTIELRLPSGEVTKKNLSDFIA 1238

RESULT 10
AAU84057
ID AAU84057 standard; Peptide; 1365 AA.
XX
AC AAU84057;
XX
DT 08-MAY-2002 (first entry)
XX
DE S. pneumoniae derived chimeric peptide, NEW30.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2..
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
```

```
XX
PR 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX
DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Example 1; Page -: 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g. caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 1365 AA;
Query Match 77.7%; Score 807; DB 23; Length 1365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 233 NMQPSQLSYSTASDNNTQSVAKGSTSKPANKSENLOLLKELDYDSPAQRYSDDGLVF 292
Db 559 NMQPSQLSYSTASDNNTQSVAKGSTSKPANKSENLOLLKELDYDSPAQRYSDDGLVF 618
Qy 293 DPAKIIISRTNGVAIPHGDHYHFIYPSKLSALEEKIARMVPISGTGSTVSTNAKPNEVVS 352
Db 619 DPAKIIISRTNGVAIPHGDHYHFIYPSKLSALEEKIARMVPISGTGSTVSTNAKPNEVVS 678
Qy 353 SLGSLSNPSSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGDHFIYIPKSNQIGQP 412
Db 679 SLGSLSNPSSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGDHFIYIPKSNQIGQP 738
Qy 413 TLPNNSLATPSPLPINPGTISHEKHEEDGYGFDANRIIAEDSGFVMSHGDNHNYFFKDD 472
Db 739 TLPNNSLATPSPLPINPGTISHEKHEEDGYGFDANRIIAEDSGFVMSHGDNHNYFFKDD 798
Qy 473 LTEEQIKAAQKHLEEVKTSINGLSDLSHSDQYPGNAKEMKDLKKIEKTAGIMKQYGV 532
Db 799 LTEEQIKAAQKHLEEVKTSINGLSDLSHSDQYPGNAKEMKDLKKIEKTAGIMKQYGV 858
Qy 533 KRESIVVVKENAIYIPHGDDHHDADIDEHKPVGIGHSNHYELFKPEEGVAKKEGNKYV 592
Db 859 KRESIVVVKENAIYIPHGDDHHDADIDEHKPVGIGHSNHYELFKPEEGVAKKEGNKYV 918
Qy 593 TGEELTNVNVLLKNSTFNNQFTLANGQKRVSFSPFPELEKLGINKLVKLITPDGKYLE 652
Db 919 TGEELTNVNVLLKNSTFNNQFTLANGQKRVSFSPFPELEKLGINKLVKLITPDGKYLE 978
```


QY 653 KVSQVFEQGVGNIANFELDPYLPQGTFFKTIASKDYPEVSDGTFTVPTSLAYKVASQ 712
 DB 979 KVSQVFEQGVGNIANFELDPYLPQGTFFKTIASKDYPEVSDGTFTVPTSLAYKVASQ 1038
 QY 713 TIFYPFHAGDYLVRNPFQFAVPKGTDALVRVDFEHGNAYLENNYKVGEIKLPIPKLNQ 772
 DB 1039 TIFYPFHAGDYLVRNPFQFAVPKGTDALVRVDFEHGNAYLENNYKVGEIKLPIPKLNQ 1098
 QY 773 TTRAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKRPSILPQFKRKAQENSKLDE 832
 DB 1099 TTRAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKRPSILPQFKRKAQENSKLDE 1158
 QY 833 KVEBPKTSEKVEKEKLSGTGNSLTENSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMD 892
 DB 1159 KVEBPKTSEKVEKEKLSGTGNSLTENSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMD 1218
 QY 893 GTIELYLPSPGEVIRKKNMADFTGEAPQNGENKPSNGKSVSTGTVENOPTENKPADSLPEA 952
 DB 1219 GTIELYLPSPGEVIRKKNMADFTGEAPQNGENKPSNGKSVSTGTVENOPTENKPADSLPEA 1278
 QY 953 PNEKPVKPSNTDNGMLNPEGVSGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1012
 DB 1279 PNEKPVKPSNTDNGMLNPEGVSGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1338
 QY 1013 IFNMDGTIELRLPSGEVIRKKNLSDFIA 1039
 DB 1339 IFNMDGTIELRLPSGEVIRKKNLSDFIA 1365

RESULT 11
 AAU84053
 ID AAU84053 standard; Peptide: 1378 AA.
 XX AC AAU84053;
 XX DT 08-MAY-2002 (first entry)
 XX DE S. pneumoniae derived chimeric peptide, .NEW26.
 XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 XX KW pneumonia; streptococcal bacterial infection; mutant; mutein;
 XX KW BVH-11-2.
 XX OS Streptococcus pneumoniae.
 XX OS Synthetic.
 XX PN WO2001198334-A2.
 XX PD 27-DEC-2001.
 XX PF 19-JUN-2001; 2001WO-CA00909.
 XX PR 20-JUN-2000; 2000US-212683P.
 XX PA (SHIR-) SHIRE BIOCHEM INC.
 XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 XX DR
 XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -
 XX Example 1; Page -; 113pp; English.
 XX PS
 XX CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3. BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX Sequence 1378 AA;

Query Match 77.7%; Score 807; DB 23; Length 1378;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NMQPSQLSYSTASDNNQTSQVAKGSTSKPANKSENQSLKELYDPSAQRYSQGLVLF 292
 DB 572 NMQPSQLSYSTASDNNQTSQVAKGSTSKPANKSENQSLKELYDPSAQRYSQGLVLF 631
 QY 293 DPAKIIISRTENGVAIPHGDHYHFIPYKLSALEEKLARMPISGTGSTVSTNAKPNEVVS 352
 DB 632 DPAKIIISRTENGVAIPHGDHYHFIPYKLSALEEKLARMPISGTGSTVSTNAKPNEVVS 691
 QY 353 SLGSLSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFFHYPKSNIGQP 412
 DB 692 SLGSLSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFFHYPKSNIGQP 751
 QY 413 TLPNNSLATPSPSLPINPGTSHKHEEDGYGFANRTIAEDESQGVMSHGDHNYHFFKND 472
 DB 752 TLPNNSLATPSPSLPINPGTSHKHEEDGYGFANRTIAEDESQGVMSHGDHNYHFFKND 811
 QY 473 LTESQIKAAQKHLEEVKTSNGLDLSSSHEDQYFGNAKEMKDLDDKITEEKIAGIMQYGV 532
 DB 812 LTESQIKAAQKHLEEVKTSNGLDLSSSHEDQYFGNAKEMKDLDDKITEEKIAGIMQYGV 871
 QY 533 KRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKEEGVAKKEGNKV 592
 DB 872 KRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKEEGVAKKEGNKV 931
 QY 593 TGEELTNVWVLLKNSFTNNQNTLANGQKRVSFSPPELEKKGGINMLVKLITPDGKVL 652
 DB 932 TGEELTNVWVLLKNSFTNNQNTLANGQKRVSFSPPELEKKGGINMLVKLITPDGKVL 991
 QY 653 KVSQVFEQGVGNIANFELDPYLPQGTFFKTIASKDYPEVSDGTFTVPTSLAYKVASQ 712
 DB 992 KVSQVFEQGVGNIANFELDPYLPQGTFFKTIASKDYPEVSDGTFTVPTSLAYKVASQ 1051
 QY 713 TIFYPFHAGDYLVRNPFQFAVPKGTDALVRVDFEHGNAYLENNYKVGEIKLPIPKLNQ 772
 DB 1052 TIFYPFHAGDYLVRNPFQFAVPKGTDALVRVDFEHGNAYLENNYKVGEIKLPIPKLNQ 1111
 QY 773 TTRAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKRPSILPQFKRKAQENSKLDE 832
 DB 1112 TTRAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKRPSILPQFKRKAQENSKLDE 1171
 QY 833 KVEBPKTSEKVEKEKLSGTGNSLTENSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMD 892
 DB 1172 KVEBPKTSEKVEKEKLSGTGNSLTENSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMD 1231
 QY 893 GTIELYLPSPGEVIRKKNMADFTGEAPQNGENKPSNGKSVSTGTVENOPTENKPADSLPEA 952
 DB 1232 GTIELYLPSPGEVIRKKNMADFTGEAPQNGENKPSNGKSVSTGTVENOPTENKPADSLPEA 1291
 QY 953 PNEKPVKPSNTDNGMLNPEGVSGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1012
 DB 1291 PNEKPVKPSNTDNGMLNPEGVSGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1351

Db 1292 PNKPKPENSTDNGLNPEGNVSGMDPLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1351
 QY 1013 IFNMDGTIELRLPSGEVIRKKNLSDFIA 1039
 Db 1352 IFNMDGTIELRLPSGEVIRKKNLSDFIA 1378

RESULT 12
 AAB12744
 ID AAB12744 standard; Protein; 780 AA.
 AC AAB12744;
 XX
 DT 21-NOV-2000 (first entry)
 DE Streptococcus pneumoniae NEW15 protein antigen SEQ ID NO:78.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA01218.
 XX
 PR 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 DR WPI; 2000-452397/39.
 XX
 Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteraemia and/or pneumonia -
 Claim 18; Fig 43; 106pp; English.

The present invention describes nucleic acids (I) encoding protein
 antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae NEW15 protein antigen.
 XX
 SQ Sequence 780 AA;

Query Match 75.1%; Score 780; DB 21; Length 780;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 780; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CAYALNHRSQENKNNRYSYVDGSSQSKSENLPDQVSOKEGTOAEQIVIKITDQGYV 80
 Db 1 CAYALNHRSQENKNNRYSYVDGSSQSKSENLPDQVSOKEGTOAEQIVIKITDQGYV 60
 QY 81 TSHGDHYHYNGKVPYDALFSEELMKDPNYOLKDADIVNEVKGGYIIKVDGKYVYLKD 140
 Db 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYOLKDADIVNEVKGGYIIKVDGKYVYLKD 120
 QY 141 AAHADNVRTKDEINPKQEHVADNEKVNNAVAVASQGRYTTNDGYVFNPAIDITDGTNA 200
 Db 121 AAHADNVRTKDEINPKQEHVADNEKVNNAVAVASQGRYTTNDGYVFNPAIDITDGTNA 180
 QY 201 YTVPHGGHYHYTPKSDLSASELAATAKAHLAGKNMOPSOLESSTASDNTQSVAKGSTSK 260
 Db 181 YTVPHGGHYHYTPKSDLSASELAATAKAHLAGKNMOPSOLESSTASDNTQSVAKGSTSK 240

QY 261 PANKSENLOSLLKELYDSPSAORYSESDGLVDFPAKIIISRTPNGVAIPHGDHVFIFYSK 320
 Db 241 PANKSENLOSLLKELYDSPSAORYSESDGLVDFPAKIIISRTPNGVAIPHGDHVFIFYSK 300
 QY 321 LSALEEKIARMPVPSGTGSTVSTNAKPNNEVSSLSGSSNPPSSLTTSKELSSASDGYIFN 380
 Db 301 LSALEEKIARMPVPSGTGSTVSTNAKPNNEVSSLSGSSNPPSSLTTSKELSSASDGYIFN 360
 QY 381 PKDIVEETATAYIVRHGDHPHYIPKSNQIGOPTLPNKSLATPSPSLPINPGTSHEKHEED 440
 Db 361 PKDIVEETATAYIVRHGDHPHYIPKSNQIGOPTLPNKSLATPSPSLPINPGTSHEKHEED 420
 QY 441 GYGFEDANRIIAEDESFGVMSHGDHNYFFKKDLTEEQIKAAQKHEEVKTSHGDLSS 500
 Db 421 GYGFEDANRIIAEDESFGVMSHGDHNYFFKKDLTEEQIKAAQKHEEVKTSHGDLSS 480
 QY 501 HEQDYPGNKEMKDLDDKIEKTIAGIMKQYGVKRESIVVWNEKNAILIYPHGDHHDADPID 560
 Db 481 HEQDYPGNKEMKDLDDKIEKTIAGIMKQYGVKRESIVVWNEKNAILIYPHGDHHDADPID 540
 QY 561 EHKPVGIGHSHSNYELFKPEGVAKKEGKVTYTGEEELTNVNNLLKNSTFNQNFLLANGQ 620
 Db 541 EHKPVGIGHSHSNYELFKPEGVAKKEGKVTYTGEEELTNVNNLLKNSTFNQNFLLANGQ 600
 QY 621 KRVSFSPPELEKKGILNMLVKLITPDGKVLKVGKVGEGVGNIANFELDQPYLPQGT 680
 Db 601 KRVSFSPPELEKKGILNMLVKLITPDGKVLKVGKVGEGVGNIANFELDQPYLPQGT 660
 QY 681 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDAL 740
 Db 661 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDAL 720
 QY 741 VRVDFEHGNAYLENNYKVGKIKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 800
 Db 721 VRVDFEHGNAYLENNYKVGKIKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 780

RESULT 13
 AAB12748
 ID AAB12748 standard; Protein; 1019 AA.
 XX
 AC AAB12748;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain WU2 BVH-3 protein antigen.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA01218.
 XX
 PR 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 DR WPI; 2000-452397/39.
 XX
 Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteraemia and/or pneumonia -
 Disclosure; Fig 11; 106pp; English.

Db	121	AAHADNVRTDEINRQKQEHVKNEKVNVAVARSGRYTTNDGVVFNPAIIIDTGN	180	OS	Streptococcus pneumoniae.		
Qy	201	YIVPHGGHYHYIPKSDLSASELAHAAGKAGKNNQPSQSSYSTASDNNTQSVAKGSTK	260	XX	WO200039299-A2.		
Db	181	YIVPHRGHYHYIPKSDLSASELAHAAGKAGKNNQPSQSSYSTASDNNTQSVAKGSTK	240	XX	06-JUL-2000.		
Qy	261	PANKSENLOQLLKYDPSAQRYSSDGLVFPDPAKIIISRTPNCAIIPHGDHYHFIPIYSK	320	PF	20-DEC-1999; 99WO-CA01218.		
Db	241	PANKSENLOQLLKYDPSAQRYSSDGLVFPDPAKIIISRTPNCAIIPHGDHYHFIPIYSK	300	XX	23-DEC-1998; 98US-0113800.		
Qy	321	LSALEEKIARMVPISGTGVSTNAKPNNEVVSGLSSNPSSLTTSKELSSASDGYIFN	380	PA	(BIOC-) BIOCHEM PHARMA INC.		
Db	301	LSALEEKIARMVPISGTGVSTNAKPNNEVVSGLSSNPSSLTTSKELSSASDGYIFN	360	PI	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;		
Qy	381	PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPINPGTSHKHEED	440	XX	WPI; 2000-452397/39.		
Db	361	PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPINPGTSHKHEED	420	PT	Streptococcal antigens useful for vaccinating against e.g. meningitis,		
Qy	441	GYGFDANRIIAEDSGFVMSHGDNHYFFKKDLTEEQIKAAQKHLEEVKTSNGLDLSL	500	PS	otitis media, bacteraemia and/or pneumonia -		
Db	421	GYGFDANRIIAEDSGFVMSHGDNHYFFKKDLTEEQIKAAQKHLEEVKTSNGLDLSL	480	XX	Disclosure; Fig 11; 106pp; English.		
Qy	501	HEQDYPGNAKEMKDLKKIEKLAGIMQYGVKRESIVVKNKEKNALIIYPHGDHHDPI	560	CC	The present invention describes nucleic acids (I) encoding protein		
Db	481	HEQDYPGNAKEMKDLKKIEKLAGIMQYGVKRESIVVKNKEKNALIIYPHGDHHDPI	540	CC	antigens (II) from Streptococcus pneumoniae. The protein antigens		
Qy	561	EHRPVGIGHSNVELFKPEGVAKGEGKVVYTGEEITNVNLLKSTNNQFTLIANGQ	620	CC	have bactericidal activity. The nucleic acids, encoding the proteins		
Db	541	EHRPVGIGHSNVELFKPEGVAKGEGKVVYTGEEITNVNLLKSTNNQFTLIANGQ	600	CC	antigens, may be used for the recombinant production of the proteins		
Qy	621	KRVSEFPPPLEKLGINMLVKLITPDGKLVKVSQKVFGEVGNIANFELDQYLPQGT	680	CC	they encode. The protein antigens may then be used as vaccines for the		
Db	601	KRVSEFPPPLEKLGINMLVKLITPDGKLVKVSQKVFGEVGNIANFELDQYLPQGT	560	CC	prevention and treatment of Streptococcal infections in mammals		
Qy	681	KYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTIYLRVNPQFAPVKGT	740	CC	(especially humans) which result in, e.g. meningitis, otitis media,		
Db	661	KYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTIYLRVNPQFAPVKGT	720	CC	bacteraemia and/or pneumonia. The present sequence represents a		
Qy	741	VRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRAGNKIPVTFMANAYLDNQSYIYE	800	CC	S. pneumoniae BVH-3 protein antigen, from the present invention.		
Db	721	VRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRAGNKIPVTFMANAYLDNQSYIYE	780	XX	Sequence 1019 AA.		
Qy	801	VPILEKENQDKPSILPOFRKNAQNSKLDKVEBPCKSEKVEKEKLSGTNSNSTL	860	Query Match	68.6%; Score 713; DB 21; Length 1019;		
Db	781	VPILEKENQDKPSILPOFRKNAQNSKLDKVEBPCKSEKVEKEKLSGTNSNSTL	840	Best Local Similarity	99.7%; Pred. No. 0;		
Qy	861	EEVPTDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSGEVIKKNMADFTGEAPGN	920	Matches 1013; Conservative	0; Mismatches 3; Indels 0; Gaps .0;		
Db	841	EEVPTDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSGEVIKKNMADFTGEAPGN	900	Qy	21	CAYALNQRHSEKNNRVSYDGSQSSQKSENLTDPQVQKQEGIAQPIVIKITDQGYV	80
Qy	921	GENKPSNGKVTGTVENQNTENKPADSLPEAPNEKVPKPNSTDNGLNPNEGVSDPM	980	Db	1	CAYALNQRHSEKNNRVSYDGSQSSQKSENLTDPQVQKQEGIAQPIVIKITDQGYV	60
Db	901	GENKPSNGKVTGTVENQNTENKPADSLPEAPNEKVPKPNSTDNGLNPNEGVSDPM	960	Qy	81	TSHGDIHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGYIHKVDGKYVYVLD	140
Qy	981	LDPALEAPAVDPQEKLEKFTASYGIGLDSVIFNMDGTIELRPSGEVIKKNLSD	1036	Db	61	TSHGDIHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGYIHKVDGKYVYVLD	120
Db	961	LDPALEAPAVDPQEKLEKFTASYGIGLDSVIFNMDGTIELRPSGEVIKKNLSD	1016	Qy	141	AAHADNVRTKDELNRQKQEHVKNEKVNVAVARSGRYTTNDGVVFNPAIIIDTGN	200
RESULT 15				Db	121	AAHADNVRTKDELNRQKQEHVKNEKVNVAVARSGRYTTNDGVVFNPAIIIDTGN	180
AA12753				Qy	201	YIVPHGGHYHYIPKSDLSASELAHAAGKAGKNNQPSQSSYSTASDNNTQSVAKGSTK	260
XX	AA12753;			Db	181	YIVPHRGHYHYIPKSDLSASELAHAAGKAGKNNQPSQSSYSTASDNNTQSVAKGSTK	240
AC	AA12753;			Qy	261	PANKSENLOQLLKYDPSAQRYSSDGLVFPDPAKIIISRTPNCAIIPHGDHYHFIPIYSK	320
XX				Db	241	PANKSENLOQLLKYDPSAQRYSSDGLVFPDPAKIIISRTPNCAIIPHGDHYHFIPIYSK	300
XX				Qy	321	LSALEEKIARMVPISGTGVSTNAKPNNEVVSGLSSNPSSLTTSKELSSASDGYIFN	380
XX				Db	301	LSALEEKIARMVPISGTGVSTNAKPNNEVVSGLSSNPSSLTTSKELSSASDGYIFN	360
XX				Qy	381	PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPINPGTSHKHEED	440
XX				Db	361	PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPINPGTSHKHEED	420
XX				Qy	441	GYGFDANRIIAEDSGFVMSHGDNHYFFKKDLTEEQIKAAQKHLEEVKTSNGLDLSL	500
XX				Db	421	GYGFDANRIIAEDSGFVMSHGDNHYFFKKDLTEEQIKAAQKHLEEVKTSNGLDLSL	480
XX				Qy	501	HEQDYPGNAKEMKDLKKIEKLAGIMQYGVKRESIVVKNKEKNALIIYPHGDHHDPI	560
XX				Db	481	HEQDYPGNAKEMKDLKKIEKLAGIMQYGVKRESIVVKNKEKNALIIYPHGDHHDPI	540

561 EHKPVGIGHSNVELFKPEGVAKKGNKVTGEEELTNVLLKNSTFNNQFTLANGQ 620
541 EHKPVGIGHSNVELFKPEGVAKKGNKVTGEEELTNVLLKNSTFNNQFTLANGQ 600
621 KRVSFSPPELEKLGIMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDPYLPQGT 680
601 KRVSFSPPELEKLGIMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDPYLPQGT 660
681 EKYTIASKDYPEVSYDGTFTVPSLAKYKMASQITFYPFHAGDYLVRNPOFAPVKGTDAL 740
661 EKYTIASKDYPEVSYDGTFTVPSLAKYKMASQITFYPFHAGDYLVRNPOFAPVKGTDAL 720
741 VRVDEFGHAGNLYNNKVGGEIKLPIPKLQGTTRTAGNKIPVTFEMANAYLDNQSYIVE 800
721 VRVDEFGHAGNLYNNKVGGEIKLPIPKLQGTTRTAGNKIPVTFEMANAYLDNQSYIVE 780
801 VPILKENQTDKPSILPOFRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLS 860
781 VPILKENQTDKPSILPOFRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLS 840
861 BEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVYKKNMADFTGEAPQGN 920
841 BEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVYKKNMADFTGEAPQGN 900
921 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVKPENSTDNGLNPEGVNGSDPM 980
901 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVKPENSTDNGLNPEGVNGSDPM 960
981 LDPALAEAPADVPQOEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVYKKNLSD 1036
961 LDPALAEAPADVPQOEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVYKKNLSD 1016

RESULT 16
AAU84092
ID AAU84092 standard; Peptide: 679 AA.
AC AAU84092;
DT 08-MAY-2002 (first entry)
XX Truncated variant of S. pneumoniae BVH-3, NEW15.
DE BVH-3; BVH-11: vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutain.
XX Streptococcus pneumoniae.
OS WO200198334-A2.
PN 27-DEC-2001.
PD 19-JUN-2001; 2001WO-CA00908.
XX 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI; 2002-122272/16.
DR New streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia -
XX Example 1; Page -; 113pp; English.
PS The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3. BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of

meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noardiae or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX Sequence 679 AA;
SQ

Query Match 65.4%; Score 679; DB 23; Length 679;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 679; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CAYALNQHSQENKDNRYSDGSSQSSOKSENLTDPQVSKQEGIAEQIVIKITDQGVY 80
DB 1 CAYALNQHSQENKDNRYSDGSSQSSOKSENLTDPQVSKQEGIAEQIVIKITDQGVY 60
QY 81 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADIIVNEKGGYIIKVDGKYVYLKD 140
DB 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADIIVNEKGGYIIKVDGKYVYLKD 120
QY 141 AAHADNVRTKDEINRQKQEHVKNVSNVAVARQGRYTTNDGVYFNPADIIETGNA 200
DB 121 AAHADNVRTKDEINRQKQEHVKNVSNVAVARQGRYTTNDGVYFNPADIIETGNA 180
QY 201 YIVPHGCHYHYIPKSDLSASELAARAHLAGKNNQPSQLSYSTASDNNNTQSVAKGSTK 260
DB 181 YIVPHGCHYHYIPKSDLSASELAARAHLAGKNNQPSQLSYSTASDNNNTQSVAKGSTK 240
QY 261 PANKSENQLSKLXYDSPAQRYSESDGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYSK 320
DB 241 PANKSENQLSKLXYDSPAQRYSESDGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYSK 300
QY 321 LSALFEKIAKRWPTSGTGVSTVNAKNEVYSSIGLSLSSNPSSLTTSKELSSASDGYIFN 380
DB 301 LSALFEKIAKRWPTSGTGVSTVNAKNEVYSSIGLSLSSNPSSLTTSKELSSASDGYIFN 360
QY 381 PKDIVEETATAYIVRHGDHFHYIPKSNQIQGPTLPNNSLATPSPSLPINFPGTSHEKHEED 440
DB 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIQGPTLPNNSLATPSPSLPINFPGTSHEKHEED 420
QY 441 GYGFDANRIIAEDESSEGVMSHGDHNYFFKFKDLTEEQIKAAQKHLEVKTSNGLDLSLS 500
DB 421 GYGFDANRIIAEDESSEGVMSHGDHNYFFKFKDLTEEQIKAAQKHLEVKTSNGLDLSLS 480
QY 501 HEODYPGNNAKEMKDLKKIEKLAGIMKQYGVKRESIVVNAKERNAIYPHGDHHDHAPID 560
DB 481 HEODYPGNNAKEMKDLKKIEKLAGIMKQYGVKRESIVVNAKERNAIYPHGDHHDHAPID 540
QY 561 EHKPVGIGHSNVELFKPEGVAKKGNKVTGEEELTNVLLKNSTFNNQFTLANGQ 620
DB 541 EHKPVGIGHSNVELFKPEGVAKKGNKVTGEEELTNVLLKNSTFNNQFTLANGQ 600
QY 621 KRVSFSPPELEKLGIMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDPYLPQGT 680
DB 601 KRVSFSPPELEKLGIMLVKLITPDGKLVLEKVGKVGEGVGNIANFELDPYLPQGT 660
QY 681 FKTYIASKDYPEVSYDGTFT 699
DB 661 FKTYIASKDYPEVSYDGTFT 679

RESULT 17

AAU83996
ID AAU83996 standard; Peptide; 644 AA.

XX AC AAU83996;

XX DT 08-MAY-2002 (first entry)

XX DE Truncated variant of S. pneumoniae BVH-11, NEW21.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX KW pneumonia; streptococcal bacterial infection; mutant; mutein.

XX OS Streptococcus pneumoniae.

XX OS Synthetic.

XX PN WO200198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.

XX PR 20-JUN-2000; 2000US-212683P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX DR WPI; 2002-122272/16.

XX CC New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX CC epitope-bearing polypeptides, useful as vaccine components for treating
XX CC or preventing streptococcal infections such as otitis media,
XX CC meningitis, and bacteraemia

XX PS Example 1; Page 7; 113pp; English.

XX CC The invention describes an isolated polypeptide (I) with 70-90%
XX CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX CC comprising (I) is useful for therapeutic or prophylactic treatment of
XX CC meningitis, otitis media, bacteraemia or pneumonia infection in an
XX CC individual susceptible to these disorders. (II) is also useful for
XX CC therapeutic or prophylactic treatment of any streptococcal bacterial
XX CC infection (e.g., caused by Streptococcus pneumoniae, group A
XX CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocardia or
XX CC Staphylococcus aureus) in an individual susceptible to the infection.
XX CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX CC techniques. The Streptococcus polypeptides are useful in a diagnostic
XX CC test for S. pneumoniae infection. (III) is useful for designing DNA
XX CC probes for use in detecting the presence of Streptococcus in a biological
XX CC sample suspected of containing the bacteria. The DNA probes may also be
XX CC used for detecting circulating S. pneumonia nucleic acid in a sample for
XX CC diagnosing streptococcal infections. This sequence represents a truncate
XX CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
XX CC described in the method of the invention.
XX CC Note: This sequence does not appear in the specification but has
XX CC been created according to information given in the invention.

XX Sequence 644 AA;

Query Match 62.0%; Score 644; DB 23; Length 644;

Best Local Similarity 100.0%; Pred. NO. 0;

Matches 644; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 396 HCDHPHYTPKSNQIQPTLPNNLSATPSLPINPGTSHKHEEDGCFDANRIIADES 455

DB 1 HGDHPHYTPKSNQIQPTLPNNLSATPSLPINPGTSHKHEEDGCFDANRIIADES 60

OY 456 GFVMSHGDNHFFKDLTEBQIKAAQKHLVEEVKTSINGLDSLSHEDQYPGNAKEMKDL 515

DB 61 GFVMSHGDNHFFKDLTEBQIKAAQKHLVEEVKTSINGLDSLSHEDQYPGNAKEMKDL 120

OY 516 DKKEEKTAGIMKOYGVKRESIVVNVKKNALIIYPHGHHADPIDKHKPVGIGHSHSNYE 575

DB 121 DKKEEKTAGIMKOYGVKRESIVVNVKKNALIIYPHGHHADPIDKHKPVGIGHSHSNYE 180

OY 576 LFKPEEGVAKKREGNKVYTGELTNVNVLLKNSTFNNOFTLANGOKRVVSFSPPELEKKL 635

DB 181 LFKPEEGVAKKREGNKVYTGELTNVNVLLKNSTFNNOFTLANGOKRVVSFSPPELEKKL 240

OY 636 GINMLVLITPDGKLVKSGVGEVGVGNIANFELDQPYLPQGTFFKYTIASKDYPEVSY 695

DB 241 GINMLVLITPDGKLVKSGVGEVGVGNIANFELDQPYLPQGTFFKYTIASKDYPEVSY 300

OY 696 DGTFTVPTSLAYKMASOTIFYPHAGDTYLRVNPQFAPVPGKTDALVRVDFEFGHAYLEN 755

DB 301 DGTFTVPTSLAYKMASOTIFYPHAGDTYLRVNPQFAPVPGKTDALVRVDFEFGHAYLEN 350

OY 756 NYKGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTIYIVEVPILEKENQTKPSI 815

DB 361 NYKGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTIYIVEVPILEKENQTKPSI 420

OY 815 LPQFKRNKAQENSKLDEKVEEPTSEKVEKLEKSETGNSTNSLTLEVPVTDVQEKVAK 875

DB 421 LPQFKRNKAQENSKLDEKVEEPTSEKVEKLEKSETGNSTNSLTLEVPVTDVQEKVAK 480

OY 876 FAESYGMKLENVLFNMDGTIELYLPISGEVKKKNMADFTGEAPOGNGENKPSSENGKYSTGT 935

DB 481 FAESYGMKLENVLFNMDGTIELYLPISGEVKKKNMADFTGEAPOGNGENKPSSENGKYSTGT 540

OY 936 VENOPTENKFSADSLPEAPNEKPKVPENSTNGMLNPEGVNGSDPMLDPALEAPAVDPVQ 995

DB 541 VENOPTENKFSADSLPEAPNEKPKVPENSTNGMLNPEGVNGSDPMLDPALEAPAVDPVQ 600

OY 996 EKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 1039

DB 601 EKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 644

RESULT 18

AAU83998

ID AAU83998 standard; Peptide; 632 AA.

XX AC AAU83998;

XX DT 08-MAY-2002 (first entry)

XX DE Truncated variant of S. pneumoniae BVH-11, NEW40.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX KW pneumonia; streptococcal bacterial infection; mutant; mutein.

XX OS Streptococcus pneumoniae.

XX OS Synthetic.

XX PN WO200198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.

XX PR 20-JUN-2000; 2000US-212683P.

XX PI (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX DR WPI; 2002-122272/16.

XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia

XX	Example 1; Page -; 113pp; English.	
PS	The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus or Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.	
CC	Note: This sequence does not appear in the specification but has been created according to information given in the invention.	
XX	Sequence 632 AA:	
SQ	Query Match 60.8%; Score 632; DB 23; Length 632; Best Local Similarity 100.0%; Pred. No. 0; Matches 632; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	408 QIGQPTLPNNLSATSPSLIPNGTSHKEHEEDGYGCFDANRIIAEDSGFVMSHGDNHY 467	
DB	1 QIGQPTLPNNLSATSPSLIPNGTSHKEHEEDGYGCFDANRIIAEDSGFVMSHGDNHY 60	
QY	468 FFKKDLTEQIAAQAQKHLSEVTSINGLDSLSHQDTPGNAKEMKDLKKEEKIAGIM 527	
DB	61 FFKKDLTEQIAAQAQKHLSEVTSINGLDSLSHQDTPGNAKEMKDLKKEEKIAGIM 120	
QY	528 KQYGVKRESIVNKENAIYIPGHDDHADPIDEHKPGVIGSHSNYELFKPEEGVAKKE 587	
DB	121 KQYGVKRESIVNKENAIYIPGHDDHADPIDEHKPGVIGSHSNYELFKPEEGVAKKE 180	
QY	588 GKNVYTGEBLTNVNLLKNTNNQNTLIANGOKRVSFSPPELEKKGINMLVLIITPD 647	
DB	181 GKNVYTGEBLTNVNLLKNTNNQNTLIANGOKRVSFSPPELEKKGINMLVLIITPD 240	
QY	648 GKVLKVGKGVGEGVNTANFELDQYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLAY 707	
DB	241 GKVLKVGKGVGEGVNTANFELDQYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLAY 300	
QY	708 KMASQTIFFPFFHAGDTYLRVNFQFVAPKGTDALVRVDFEFGNAYLNNYKVGEEKLPITP 767	
DB	301 KMASQTIFFPFFHAGDTYLRVNFQFVAPKGTDALVRVDFEFGNAYLNNYKVGEEKLPITP 360	
QY	768 KLNQGTTRAGNKIPVTFMANAYLDNQSYIYVEVPILEKXENQTDPSIILPOFKRNKAQEN 827	
DB	361 KLNQGTTRAGNKIPVTFMANAYLDNQSYIYVEVPILEKXENQTDPSIILPOFKRNKAQEN 420	
QY	828 SKLDEKVEEPKTSKVEKEKLSGTNSTLSNLTSEVPTVDPQVKVAKFAESYGMKLENY 887	
DB	421 SKLDEKVEEPKTSKVEKEKLSGTNSTLSNLTSEVPTVDPQVKVAKFAESYGMKLENY 480	
QY	888 LFNMDGTIELYLPSPGSEVVKKNMADFTGEAPQNGENKPSNGKVSTGTVENOPTENKPAD 947	
DB	481 LFNMDGTIELYLPSPGSEVVKKNMADFTGEAPQNGENKPSNGKVSTGTVENOPTENKPAD 540	
QY	948 SLPEAPNEKPVKPNSTNDGMLNPEGNVGSDPMLDPALEAPAVDPQVKLEKFTASYGL 1007	
DB	541 SLPEAPNEKPVKPNSTNDGMLNPEGNVGSDPMLDPALEAPAVDPQVKLEKFTASYGL 600	
QY	1008 GLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 1039	

DB	601 GLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 632	
RESULT 19		
AAB12749		
ID	AAB12749 standard; Protein; 1019 AA.	
XX	AAB12749:	
DT	21-NOV-2000 (first entry)	
DE	Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.	
KW	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia; otitis media; pneumonia; immunisation; bactericidal.	
OS	Streptococcus pneumoniae.	
PN	WO200039299-A2.	
PD	06-JUL-2000.	
PF	20-DEC-1999; 99WO-CA01218.	
PR	23-DEC-1998; 98US-0113800.	
XX	(BIOC-) BIOCHEM PHARMA INC.	
XX	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;	
DR	WPI: 2000-452397/39.	
PT	Streptococcal antigens useful for vaccinating against e.g. meningitis, otitis media, bacteraemia and/or pneumonia	
PS	Disclosure: Fig 11: 106pp; English.	
CC	The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents a S. pneumoniae BVH-3 protein antigen, from the present invention.	
SQ	Sequence 1019 AA:	
Query Match	58.9%; Score 612; DB 21; Length 1019;	
Best Local Similarity	99.6%; Pred. No. 0;	
Matches 1012; Conservative	0; Mismatches 4; Indels 0; Gaps 0;	
QY	21 CAYALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSKQEGIAQBIKIDQGYV 80	
DB	1 CAYALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVSKQEGIAQBIKIDQGYV 60	
QY	81 TSHGDHYHYNGKVPYDALFSELLMKDPNYQLKADIVNEVKGYITKVDGKYVYVLD 140	
DB	61 TSHGDHYHYNGKVPYDALFSELLMKDPNYQLKADIVNEVKGYITKVDGKYVYVLD 120	
QY	141 AAHADNVRTKQINRQKHVKDNEKNSNVAVARSQGRYTTNDGYVFNPAIDIEDTGN 200	
DB	121 AAHADNVRTKQINRQKHVKDNEKNSNVAVARSQGRYTTNDGYVFNPAIDIEDTGN 180	
QY	201 YIVPHGGHYHYTPKSDLSASELAALAAKHAHLAGKNMOPQSLSYSSSTASDNNTOSVAKGTSK 260	
DB	181 YIVPHGGHYHYTPKSDLSASELAALAAKHAHLAGKNMOPQSLSYSSSTASDNNTOSVAKGTSK 240	
QY	261 PANKSENQSLKELYDPSAQRYSESGLYFDPKAITSRTPNGVAIPGHDHYHPIYSK 320	

Db 241 PANKSENUGSLKELYDSSAQRYSSEGLVDFDPAKIIISRTPNGVAIPHGHHYHPIPYSK 300
Qy 321 LSALEKIAIRWPIISGTGTSTNNAKPNVSSLSGSLSNPSSLTTSKELSSASDGYIFN 380
Db 301 LSALEKIAIRWPIISGTGTSTNNAKPNVSSLSGSLSNPSSLTTSKELSSASDGYIFN 360
Qy 381 PKDIVEETATYIVRHGDHFIYIPKSNQIGQPTLNNSLATPSPSLPINPTSHEKHEED 440
Db 361 PKDIVEETATYIVRHGDHFIYIPKSNQIGQPTLNNSLATPSPSLPINPTSHEKHEED 420
Qy 441 GYGFANDRIIADESGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSHNGLDLSLS 500
Db 421 GYGFANDRIIADESGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSHNGLDLSLS 480
Qy 501 HEODYPGNAKEMKDLDDKTEEKIAGIMQYGVKRESIVVYKKNIAIYPHGHHADPID 560
Db 481 HEODYPGNAKEMKDLDDKTEEKIAGIMQYGVKRESIVVYKKNIAIYPHGHHADPID 540
Qy 561 EHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGELTNVNVLLKNSFNQNFPLANGQ 620
Db 541 EHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGELTNVNVLLKNSFNQNFPLANGQ 600
Qy 621 KRVSFPPPELEKKGILNMLVLIITPDGKVLKSVGKVGEGVGNIAFELDQVLPQQT 680
Db 601 KRVSFPPPELEKKGILNMLVLIITPDGKVLKSVGKVGEGVGNIAFELDQVLPQQT 660
Qy 681 FKYTIAASKDYPEVSDGTFTVPTSLAYKMASOTIYPFHAGDTYLRVNPQFAVPKGTAL 740
Db 661 FKYTIAASKDYPEVSDGTFTVPTSLAYKMASOTIYPFHAGDTYLRVNPQFAVPKGTAL 720
Qy 741 VRYDFEHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 800
Db 721 VRYDFEHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 780
Qy 801 VPILKENGTDKPSILPQKRNKAQENSKLDEKVEEPTSEKVEKEKISETGNSTNSL 860
Db 781 VPILKENGTDKPSILPQKRNKAQENSKLDEKVEEPTSEKVEKEKISETGNSTNSL 840
Qy 861 EYPTVDVPQVEKAKPAESYGMKLENVLFNMDGTIELYLPSEGEVKKKNMADFTGEAPQGN 920
Db 841 EYPTVDVPQVEKAKPAESYGMKLENVLFNMDGTIELYLPSEGEVKKKNMADFTGEAPQGN 900
Qy 921 GENKPSGENKVTGTVENOPTENKPADSLPEAPNEKPKVPENSTNGMLNPGNVSDDPM 980
Db 901 GENKPSGENKVTGTVENOPTENKPADSLPEAPNEKPKVPENSTNGMLNPGNVSDDPM 960
Qy 981 LDPALAEAPAVDPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSD 1036
Db 961 LDPALAEAPAVDPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSD 1016

RESULT 20
AAU84002
ID AAU84002 standard; Peptide; 632 AA.
XX AAU84002;
AC AAU84002;
XX 08-MAY-2002 (first entry)
DT Truncated variant of S. pneumoniae BVH-3, NEW49.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.
XX Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
PN 27-DEC-2001.
XX 19-JUN-2001; 2001WO-CA00908.
PF
XX

PR 20-JUN-2000; 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI: 2002-122272/16.
XX
DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Claim 1; Page : 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 632 AA;
Query Match 55.6%; Score 578; DB 23; Length 632;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 452 GDHNYHFFKDLTEQTKAAQKHLEEVKTSHNGLDLSLSHEDQYPGNAKEMKDLKKIEE 521
Db 55 GDHNYHFFKDLTEQTKAAQKHLEEVKTSHNGLDLSLSHEDQYPGNAKEMKDLKKIEE 114
Qy 522 KIAGIMQYGVKRESIVVYKKNIAIYPHGHHADPIDDEKHPVIGHSNVELFKPEE 581
Db 115 KIAGIMQYGVKRESIVVYKKNIAIYPHGHHADPIDDEKHPVIGHSNVELFKPEE 174
Qy 582 GVAKKEGKVYTGELTNVNVLLKNSFNQNFPLANGQKRVSPSPFPELEKILGINLV 641
Db 175 GVAKKEGKVYTGELTNVNVLLKNSFNQNFPLANGQKRVSPSPFPELEKILGINLV 234
Qy 642 KLITPDGKVLKSVGKVGEGVGNIAFELDQVLPQQTFTKTYTIAASKDYPEVSDGFTV 701
Db 235 KLITPDGKVLKSVGKVGEGVGNIAFELDQVLPQQTFTKTYTIAASKDYPEVSDGFTV 294
Qy 702 PTLAYKMASOTIYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVEE 761
Db 295 PTLAYKMASOTIYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVEE 354
Qy 762 IKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEPILKENGTDKPSILPQFKR 821
Db 355 IKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEPILKENGTDKPSILPQFKR 414
Qy 822 NKAQENSKLDEKVEEPTSEKVEKEKISETGNSTNSLTLEEVPTVDVPQVEKAKFAESYG 881
Db 415 NKAQENSKLDEKVEEPTSEKVEKEKISETGNSTNSLTLEEVPTVDVPQVEKAKFAESYG 474
Qy 882 MKLENVLFNMDGTIELYLPSEGEVKKKNMADFTGEAPQGNENKPSGENKVTGTVENOPT 941

Db 475 MLENVLFNMDGTIELYLPSEGEVIKKNMADTGEAPOGNGENKPSSENGKSTGTVENOPT 534
Qy 942 ENKPADSLPEAPNEKPVKPNSTNGMLNPGNGVSDPMLDPALEAPAVDPVQEKLEKF 1001
Db 535 ENKPADSLPEAPNEKPVKPNSTNGMLNPGNGVSDPMLDPALEAPAVDPVQEKLEKF 594
Qy 1002 TASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 595 TASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 632

RESULT 21
AAU84003
ID AAU84003 standard; Peptide; 632 AA.
AC AAU84003;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-3, NEW50.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; muten.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI: 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Claim 1; Page -: 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 632 AA;

Query Match 55.3%; Score 575; DB 23; Length 632;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 465 NHYFFKDLTEEQTKAAQKHLEEVKTSNGLDLSLSSEHODYPGNAKEMKDKKIEKIA 524
Db 58 NHYFFKDLTEEQTKAAQKHLEEVKTSNGLDLSLSSEHODYPGNAKEMKDKKIEKIA 117
Qy 525 GIMQYGVKRESIVVNEKKNATYYPHGDHHDPIIDHKPVGIGHSHSNYELEKPEEGVA 584
Db 118 GIMQYGVKRESIVVNEKKNATYYPHGDHHDPIIDHKPVGIGHSHSNYELEKPEEGVA 177
Qy 585 KKEGNKYVTGEELTNVNVNLLKNSTFNQNTLANGQRVRSFSPPELEKLGIMLVKLI 644
Db 178 KKEGNKYVTGEELTNVNVNLLKNSTFNQNTLANGQRVRSFSPPELEKLGIMLVKLI 237
Qy 645 TPDCKVLEKVGSKVFGGVGNIANFELDQVLPQCTEKYTIASKDYDEVSYDGTFTVPTS 704
Db 238 TPDCKVLEKVGSKVFGGVGNIANFELDQVLPQCTEKYTIASKDYDEVSYDGTFTVPTS 297
Qy 705 LAYKMASQTIIFYPHAGDTYLRVNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGIEKL 764
Db 298 LAYKMASQTIIFYPHAGDTYLRVNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGIEKL 357
Qy 765 PIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEPILKKNQTDKPSILPOFKENKA 824
Db 358 PIPKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEPILKKNQTDKPSILPOFKENKA 417
Qy 825 QENSKLDEKVEEPTSEKVEKELSETGNSTNSLTSEVPTVDPVQEKVAKFAESYGMKL 884
Db 418 QENSKLDEKVEEPTSEKVEKELSETGNSTNSLTSEVPTVDPVQEKVAKFAESYGMKL 477
Qy 885 ENVLFNMDGTIELYLPSEGEVIKKNMADTGEAPOGNGENKPSSENGKSTGTVENOPTENK 944
Db 478 ENVLFNMDGTIELYLPSEGEVIKKNMADTGEAPOGNGENKPSSENGKSTGTVENOPTENK 537
Qy 945 PADSLPEAPNEKPVKPNSTNGMLNPGNGVSDPMLDPALEAPAVDPVQEKLEKFTAS 1004
Db 538 PADSLPEAPNEKPVKPNSTNGMLNPGNGVSDPMLDPALEAPAVDPVQEKLEKFTAS 597
Qy 1005 YGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 598 YGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 632

RESULT 22
AAU84013
ID AAU84013 standard; Peptide; 632 AA.
AC AAU84013;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-3, NEW63.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; muten.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX DR WPI; 2002-122272/16.

XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

XX PT epitope-bearing polypeptides, useful as vaccine components for treating

XX PT or preventing streptococcal infections such as otitis media,

XX PT meningitis, and bacteraemia

XX PS Claim 1; Page -: 113pp; English.

XX CC The invention describes an isolated polypeptide (I) with 70-90%

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or

CC Streptococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate

CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

XX SX Sequence 632 AA;

Query Match 55.3%; Score 575; DB 23; Length 632;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 NHYFFKDLTEQIAAKHLEEVKTSNGLDLSLSEODYPCGNKAKEMKLDKKEEIA 524

DB 58 NHYFFKDLTEQIAAKHLEEVKTSNGLDLSLSEODYPCGNKAKEMKLDKKEEIA 117

QY 525 GIMKQGVKRESIVNKENAIYIPHGDDHHDADIDEHKPVGIGHSHSNYELFKPEEVA 584

DB 118 GIMKQGVKRESIVNKENAIYIPHGDDHHDADIDEHKPVGIGHSHSNYELFKPEEVA 177

QY 585 KKEGNKVTGTELTNNVLLKNTFNQNTLANGOKRVSFSPPELEKKGILNMLVKLI 644

DB 178 KKEGNKVTGTELTNNVLLKNTFNQNTLANGOKRVSFSPPELEKKGILNMLVKLI 237

QY 645 TPDGKVLKSVKVGEGVGNFIANFELDQPLPGOTFKYITASKDYPEVSDGTFTVPTS 704

DB 238 TPDGKVLKSVKVGEGVGNFIANFELDQPLPGOTFKYITASKDYPEVSDGTFTVPTS 297

QY 705 LAYKASQIIFPEHAGDYLVRNPOFAPVKGTDALVRVDFEHCNAYLNNYKVGTEKL 764

DB 298 LAYKASQIIFPEHAGDYLVRNPOFAPVKGTDALVRVDFEHCNAYLNNYKVGTEKL 357

QY 765 PIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEPFIKENOTDKPSILPOFKRKA 824

DB 358 PIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEPFIKENOTDKPSILPOFKRKA 417

QY 825 QNSKLDKEVEPKTSKVEKLETSNSTLEEVPTVDPVQEKVAFPAESYGMKL 884

DB 418 QNSKLDKEVEPKTSKVEKLETSNSTLEEVPTVDPVQEKVAFPAESYGMKL 477

QY 885 ENVLNMDGTIELYLPSEGVKIKKNMADTGEAPQNGENKPSSENGKVGSTGVENQPTENK 944

DB 478 ENVLNMDGTIELYLPSEGVKIKKNMADTGEAPQNGENKPSSENGKVGSTGVENQPTENK 537

QY 945 PADSLPEAPNEKPKPENSTNGMLNPGNVGSDPMLDPALEEAPVDPVOEKLKFTAS 1004

DB 538 PADSLPEAPNEKPKPENSTNGMLNPGNVGSDPMLDPALEEAPVDPVOEKLKFTAS 597

QY 1005 YGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 1039

DB 598 YGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 632

RESULT 23

AAU84018

ID AAU84018 standard; Peptide: 626 AA.

XX AC AAU84018;

XX DT 08-MAY-2002 (first entry)

XX DE Truncated variant of S. pneumoniae BVH-3, NEW105.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX KW pneumonia; streptococcal bacterial infection; mutant; mutein.

XX OS Streptococcus pneumoniae.

XX OS Synthetic.

XX WO200198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.

XX PR 20-JUN-2000; 2000US-212683P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/16.

XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

XX PT epitope-bearing polypeptides, useful as vaccine components for treating

XX PT or preventing streptococcal infections such as otitis media,

XX PT meningitis, and bacteraemia

XX PS Claim 1; Page -: 113pp; English.

XX CC The invention describes an isolated polypeptide (I) with 70-90%

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or

CC Streptococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate

CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

XX SX Sequence 626 AA;

Query Match 55.1%; Score 573; DB 23; Length 626;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 467 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSEHQDYPGNAKEMKDLKKIEKIAGI 526
Db 54 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSEHQDYPGNAKEMKDLKKIEKIAGI 113
QY 527 MKQYGVKRESIVVNEKNAIIPYHGHHADPIDEHKVPVIGSHSNYELFKPEEGVAKK 586
Db 114 MKQYGVKRESIVVNEKNAIIPYHGHHADPIDEHKVPVIGSHSNYELFKPEEGVAKK 173
QY 587 EGNKYVTGEELTNVNLKNSFNQNTLANGQKRVFSFPPPELEKLGINMLVKLITP 646
Db 174 EGNKYVTGEELTNVNLKNSFNQNTLANGQKRVFSFPPPELEKLGINMLVKLITP 233
QY 647 DGKVLKESVKGFGVGVGNIAFELDQPLPGQTEFKYTIASKDYPEVSYDGTFTVPTSLA 706
Db 234 DGKVLKESVKGFGVGVGNIAFELDQPLPGQTEFKYTIASKDYPEVSYDGTFTVPTSLA 293
QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFDEPHGNAYLENNYKVGEIKLPI 766
Db 294 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFDEPHGNAYLENNYKVGEIKLPI 353
QY 767 PKLNGQTRTAGNKIPVTFMANAYLDNQSTIYVEVPILKEKNQTDKPSILPQFRNKAAQE 826
Db 354 PKLNGQTRTAGNKIPVTFMANAYLDNQSTIYVEVPILKEKNQTDKPSILPQFRNKAAQE 413
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVTPDPVQEKVAKFAESYGMKLEN 886
Db 414 NSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVTPDPVQEKVAKFAESYGMKLEN 473
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKYSTGTGVNQPTENKPA 946
Db 474 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKYSTGTGVNQPTENKPA 533
QY 947 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 534 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 593
QY 1007 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
Db 594 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 626

RESULT 24
AAU84014
ID AAU84014 standard; Peptide; 632 AA.
XX
AC AAU84014;
XX
XX 08-MAY-2002 (first entry)
XX
XX Truncated variant of S. pneumoniae BVH-3, NEW64.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2.
XX
XX 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
PT
```

```
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
XX Claim 1: Page 1: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX
```

Sequence 632 AA;

Query Match 55.1%; Score 573; DB 23; Length 632;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 467 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSEHQDYPGNAKEMKDLKKIEKIAGI 526
Db 60 YFFKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSEHQDYPGNAKEMKDLKKIEKIAGI 119
QY 527 MKQYGVKRESIVVNEKNAIIPYHGHHADPIDEHKVPVIGSHSNYELFKPEEGVAKK 586
Db 120 MKQYGVKRESIVVNEKNAIIPYHGHHADPIDEHKVPVIGSHSNYELFKPEEGVAKK 179
QY 587 EGNKYVTGEELTNVNLKNSFNQNTLANGQKRVFSFPPPELEKLGINMLVKLITP 646
Db 180 EGNKYVTGEELTNVNLKNSFNQNTLANGQKRVFSFPPPELEKLGINMLVKLITP 239
QY 647 DGKVLKESVKGFGVGVGNIAFELDQPLPGQTEFKYTIASKDYPEVSYDGTFTVPTSLA 706
Db 240 DGKVLKESVKGFGVGVGNIAFELDQPLPGQTEFKYTIASKDYPEVSYDGTFTVPTSLA 299
QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFDEPHGNAYLENNYKVGEIKLPI 766
Db 300 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFDEPHGNAYLENNYKVGEIKLPI 359
QY 767 PKLNGQTRTAGNKIPVTFMANAYLDNQSTIYVEVPILKEKNQTDKPSILPQFRNKAAQE 826
Db 360 PKLNGQTRTAGNKIPVTFMANAYLDNQSTIYVEVPILKEKNQTDKPSILPQFRNKAAQE 419
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVTPDPVQEKVAKFAESYGMKLEN 886
Db 420 NSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVTPDPVQEKVAKFAESYGMKLEN 479
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKYSTGTGVNQPTENKPA 946
Db 480 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKYSTGTGVNQPTENKPA 539
QY 947 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 540 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 599
QY 1007 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
Db 600 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 632
```

RESULT 25
AAU84070
ID AAU84070 standard; Peptide; 895 AA.
XX AC AAU84070;
XX DT 08-MAY-2002 (first entry)
XX DE S. pneumoniae derived chimeric peptide, VP112.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX KW pneumonia; streptococcal bacterial infection; mutant; mutain;
XX KW BVH-11-2.
XX OS Streptococcus pneumoniae.
XX OS Synthetic.
XX PN WO200198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX DR WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
XX Claim 1; Page -; 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX Sequence 895 AA;
XX
XX Query Match 55.1%; Score 573; DB 23; Length 895;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 467 YFFPKDLTEQIKAAQKHLEEVKTSNGLSDLSSSHQDPYFGNAKEMKDLDKIEKIKI 526
XX
XX 323 YFFPKDLTEQIKAAQKHLEEVKTSNGLSDLSSSHQDPYFGNAKEMKDLDKIEKIKI 382
XX

QY 527 MKQYGVRESIVVKNKRNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586
DB 383 MKQYGVRESIVVKNKRNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 442
QY 587 EGNKVYTGELTNVNNLLKNSTFNNQNTLANGOKRVSFSPPELEKKGGINMLVLIITP 646
DB 443 EGNKVYTGELTNVNNLLKNSTFNNQNTLANGOKRVSFSPPELEKKGGINMLVLIITP 502
QY 647 DGKYLEKVGKVFGEVGNITANFELDQYLPQGTFFKTYIASKDYPEVSYDGTFTVPTSLA 706
DB 503 DGKYLEKVGKVFGEVGNITANFELDQYLPQGTFFKTYIASKDYPEVSYDGTFTVPTSLA 562
QY 707 YKMASQTIFFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKYGEIKLPI 766
DB 563 YKMASQTIFFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKYGEIKLPI 622
QY 767 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEPPILEKENQTDKPSILPQFKRNKAQE 826
DB 623 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEPPILEKENQTDKPSILPQFKRNKAQE 682
QY 827 NSKLDKVEEPPKTSKVEKEKLSGTGNTSTNSILSEVPTVDPVOEKYAKFAESYGHKLEN 886
DB 683 NSKLDKVEEPPKTSKVEKEKLSGTGNTSTNSILSEVPTVDPVOEKYAKFAESYGHKLEN 742
QY 887 VLFNMDGTIELYLPSEGEVVKKNMADFTGEAPQNGENKPSNGKSVSTGTVENOPTENKPA 946
DB 743 VLFNMDGTIELYLPSEGEVVKKNMADFTGEAPQNGENKPSNGKSVSTGTVENOPTENKPA 802
QY 947 DSLPEAPNEKVRPENSTDNGLMNPENGVSDPMLDPALEAPAVDPVQKLEKFTASYG 1006
DB 803 DSLPEAPNEKVRPENSTDNGLMNPENGVSDPMLDPALEAPAVDPVQKLEKFTASYG 862
QY 1007 LGLDSVIFNMDGTIELRPSGEVVKKNLSDFIA 1039
DB 863 LGLDSVIFNMDGTIELRPSGEVVKKNLSDFIA 895
RESULT 26
AAU84080
ID AAU84080 standard; Peptide; 896 AA.
XX AC AAU84080;
XX DT 08-MAY-2002 (first entry)
XX DE S. pneumoniae derived chimeric peptide, VP123.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX KW pneumonia; streptococcal bacterial infection; mutant; mutain;
XX OS Streptococcus pneumoniae.
XX OS Synthetic.
XX PN WO200198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX DR WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX

```
PS Claim 1: Page -: 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 896 AA;

Query Match 55.1%; Score 573; DB 23; Length 896;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKKDLTEQIKAAQKHLEEVKTSNGHLDLSLSSHQDYPGNAKENKMDLDKIEKIAGI 526
Db 324 YFFKKDLTEQIKAAQKHLEEVKTSNGHLDLSLSSHQDYPGNAKENKMDLDKIEKIAGI 383

QY 527 MKQYGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGICGSHSNYELFKPEGVAKK 586
Db 384 MKQYGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGICGSHSNYELFKPEGVAKK 443

QY 587 EGNKYVTGELNVNLLKNSFNQNFANGQKRVSFSPPELEKLGINLWKLITP 646
Db 444 EGNKYVTGELNVNLLKNSFNQNFANGQKRVSFSPPELEKLGINLWKLITP 503

QY 647 DGKVLKSVGKVFEGVGNIANFELDPQYLPQGTFRYTIASKDYPVSDGFTVPTSLA 706
Db 504 DGKVLKSVGKVFEGVGNIANFELDPQYLPQGTFRYTIASKDYPVSDGFTVPTSLA 563

QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDLALVRVDFEFGHGNAYLENNYKVGEIKLPI 766
Db 564 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDLALVRVDFEFGHGNAYLENNYKVGEIKLPI 623

QY 767 PKLNQGTTRTAGNKIEVTFMANAYLDNQSTYIVVEPILKENQTDKPSILPQFKNKAE 826
Db 624 PKLNQGTTRTAGNKIEVTFMANAYLDNQSTYIVVEPILKENQTDKPSILPQFKNKAE 683

QY 827 NSKLDEKVEEPTSEKVEKEKLSNSETNSNSTLEEVTPDVPQEKVAFASVGMKLEN 886
Db 684 NSKLDEKVEEPTSEKVEKEKLSNSETNSNSTLEEVTPDVPQEKVAFASVGMKLEN 743

QY 887 VLFNMDGTIELYLPSEGVTKKNMADTGAPOGNGENKPSSENGKSTGVNQPTENKPA 946
Db 744 VLFNMDGTIELYLPSEGVTKKNMADTGAPOGNGENKPSSENGKSTGVNQPTENKPA 803

QY 947 DSLPEAPNEKPKPENSTNGMLNPSGNGVSDPMDLPALEAPAVDPVQEKLEKFTASVG 1006
Db 804 DSLPEAPNEKPKPENSTNGMLNPSGNGVSDPMDLPALEAPAVDPVQEKLEKFTASVG 863

QY 1007 LGLDSVIFNMDDGTIELRPSGVEIKKNLSDFIA 1039
Db 864 LGLDSVIFNMDDGTIELRPSGVEIKKNLSDFIA 896

RESULT 27
```

QY 587 EGNKYTGEEITNVNLLKNSFTNNQFTLANGOKRVSEFPPPELEKKGILGNMLKLTLP 646
 DB 449 EGNKYTGEEITNVNLLKNSFTNNQFTLANGOKRVSEFPPPELEKKGILGNMLKLTLP 508
 QY 647 DGKLVKSGVKGEGVGNITANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLA 706
 DB 509 DGKLVKSGVKGEGVGNITANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLA 568
 QY 707 YKMASQITIFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGELKLP 766
 DB 569 YKMASQITIFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGELKLP 628
 QY 767 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKNTDKPSILPOFKRKAQE 826
 DB 629 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKNTDKPSILPOFKRKAQE 688
 QY 827 NSKLDEKVEEPTSEKVEKELSETGNSNSTLEEVPTVDPQEKVAFASGKMLN 886
 DB 689 NSKLDEKVEEPTSEKVEKELSETGNSNSTLEEVPTVDPQEKVAFASGKMLN 748
 QY 887 VLFNMDGTIELYLPSEGEVIRKKNMADFTGEAPQNGENKPSENGKSTGTIVENQPTENKPA 946
 DB 749 VLFNMDGTIELYLPSEGEVIRKKNMADFTGEAPQNGENKPSENGKSTGTIVENQPTENKPA 808
 QY 947 DSLPEAPNEKVPKPNSTNGMLNPEGNVGSDPMLDPALEAPVDPVQEKLEKFTASYG 1006
 DB 809 DSLPEAPNEKVPKPNSTNGMLNPEGNVGSDPMLDPALEAPVDPVQEKLEKFTASYG 868
 QY 1007 LGLDSVIFNMDGTIELRPSGEVIRKKNLSDFIA 1039
 DB 869 LGLDSVIFNMDGTIELRPSGEVIRKKNLSDFIA 901

RESULT 28

AAU84072
 ID AAU84072 standard; Peptide: 901 AA.

AC AAU84072;
 XX

DT 08-MAY-2002 (first entry)
 XX

DE S. pneumoniae derived chimeric peptide, vpl14.
 XX

KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 .KW pneumonia; streptococcal bacterial infection; mutant; mutain;
 KW BVH-11-2.
 XX

OS Streptococcus pneumoniae.
 OS Synthetic.
 XX

FN WO200198334-A2.
 XX

XX 27-DEC-2001.
 PD

XX 19-JUN-2001; 2001WO-CA00908.
 PF

XX 20-JUN-2000; 2000US-212683P.
 PR

XX (SHIR-) SHIRE BIOCHEM INC.
 PA

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX

DR WPI; 2002-122272/16.
 XX

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -
 XX

PS Claim 1; Page -; 113pp; English.
 PS

XX The invention describes an isolated polypeptide (I) with 70-90%
 CC

CC identity to streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noecardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 CC

XX Sequence 901 AA;

Query Match 55.1%; Score 573; DB 23; Length 901;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKKDLTETESQIAAKQHLKEEVKTSUNGSLSSSHEDQYPGNAKEMKDLCKIEEKIAGI 526

DB 329 YFFKKDLTETESQIAAKQHLKEEVKTSUNGSLSSSHEDQYPGNAKEMKDLCKIEEKIAGI 388

QY 527 MKQYGVKRESIVVNKKNALIIYPGHHDHADPTDEKPKVIGHSNHYELFPEEGVAKK 586

DB 389 MKQYGVKRESIVVNKKNALIIYPGHHDHADPTDEKPKVIGHSNHYELFPEEGVAKK 448

QY 587 EGNKYTGEEITNVNLLKNSFTNNQFTLANGOKRVSEFPPPELEKKGILGNMLKLTLP 646

DB 449 EGNKYTGEEITNVNLLKNSFTNNQFTLANGOKRVSEFPPPELEKKGILGNMLKLTLP 508

QY 647 DGKLVKSGVKGEGVGNITANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLA 706

DB 509 DGKLVKSGVKGEGVGNITANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLA 568

QY 707 YKMASQITIFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGELKLP 766

DB 569 YKMASQITIFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGELKLP 628

QY 767 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKNTDKPSILPOFKRKAQE 826

DB 629 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKNTDKPSILPOFKRKAQE 688

QY 827 NSKLDEKVEEPTSEKVEKELSETGNSNSTLEEVPTVDPVQEKVAFASGKMLN 886

DB 689 NSKLDEKVEEPTSEKVEKELSETGNSNSTLEEVPTVDPVQEKVAFASGKMLN 748

QY 887 VLFNMDGTIELYLPSEGEVIRKKNMADFTGEAPQNGENKPSENGKSTGTIVENQPTENKPA 946

DB 749 VLFNMDGTIELYLPSEGEVIRKKNMADFTGEAPQNGENKPSENGKSTGTIVENQPTENKPA 808

QY 947 DSLPEAPNEKVPKPNSTNGMLNPEGNVGSDPMLDPALEAPVDPVQEKLEKFTASYG 1006

DB 809 DSLPEAPNEKVPKPNSTNGMLNPEGNVGSDPMLDPALEAPVDPVQEKLEKFTASYG 868

QY 1007 LGLDSVIFNMDGTIELRPSGEVIRKKNLSDFIA 1039

DB 869 LGLDSVIFNMDGTIELRPSGEVIRKKNLSDFIA 901

RESULT 29

AAU84074

ID AAU84074 standard; Peptide: 901 AA.

XX

AC	AAU84074;
XX	08-MAY-2002 (first entry)
DE	S. pneumoniae derived chimeric peptide, VPI16.
XX	BVH-3; BVH-11: vaccine: meningitis; otitis media; bacteraemia;
KW	pneumonia; streptococcal bacterial infection; mutant; mutein;
XX	BVH-11-2.
XX	Streptococcus pneumoniae.
OS	Synthetic.
XX	WO200198334-A2.
PN	27-DEC-2001.
XX	19-JUN-2001; 2001WO-CA00908.
PD	20-JUN-2000; 2000US-212683P.
XX	(SHIR-) SHIRE BIOCHEM INC.
PA	Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI	WPI: 2002-122272/16.
DR	New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT	epitope-bearing polypeptides, useful as vaccine components for treating
PT	or preventing streptococcal infections such as otitis media,
PT	meningitis, and bacteraemia
XX	Claim 1; page -; 113pp; English.
PS	The invention describes an isolated polypeptide (I) with 70-90%
CC	identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC	BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC	comprising (I) is useful for therapeutic or prophylactic treatment of
CC	meningitis, otitis media, bacteraemia or pneumonia infection in an
CC	individual susceptible to these disorders. (II) is also useful for
CC	therapeutic or prophylactic treatment of any streptococcal bacterial
CC	infection (e.g., caused by Streptococcus pneumoniae, group A
CC	Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC	as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
CC	Staphylococcus aureus) in an individual susceptible to the infection.
CC	A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC	techniques. The Streptococcus polypeptides are useful in a diagnostic
CC	test for S. pneumoniae infection. (III) is useful for designing DNA
CC	probes for use in detecting the presence of Streptococcus in a biological
CC	sample suspected of containing the bacteria. The DNA probes may also be
CC	used for detecting circulating S. pneumonia nucleic acid in a sample for
CC	diagnosing streptococcal infections. This sequence represents a chimeric
CC	gene created from fragments and variant fragments of Streptococcus
CC	pneumoniae genes, described in the method of the invention.
CC	Note: this sequence does not appear in the specification but has
CC	been created according to information given in the invention.
XX	Sequence 901 AA;
SQ	Query Match 55.1%; Score 573; DB 23; Length 901;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	467 YFFKDLTEQKAQAQHLEEVKTSNGLDLSLSSHEQDYFGNAKEMKDDDKITEEKIAGI 526
Db	329 YFFKDLTEQKAQAQHLEEVKTSNGLDLSLSSHEQDYFGNAKEMKDDDKITEEKIAGI 388
QY	527 MKQYGVKRESIVVNEKENAIYPGHDHHADPIDEHKPVGIHSHSNYELFPEEGVAKK 586
Db	389 MKQYGVKRESIVVNEKENAIYPGHDHHADPIDEHKPVGIHSHSNYELFPEEGVAKK 448
QY	587 EGNKYVTGBELTNVYNLLKNSTFNQNFTLANGOKRYSFSFPPELEBKUGINMLKLITP 646

Db	449	EGNKVYTGEBLTINWLLKNSFTNQNQFTLANGQKRVRSFSPFPELEKKLGLINMLVKLIITP	508
Qy	647	DGKVLKSVKVFEGVGNIANFELDQPLPGQTETKYTIASKDYPEVSYDGTFTVPTSLA	705
Db	509	DGKVLKSVKVFEGVGNIANFELDQPLPGQTETKYTIASKDYPEVSYDGTFTVPTSLA	568
Qy	707	YKMSQTIFYPFHAGDTYLRVNPQFAPVPGKTDALVRVDFDEHGNAYLENNYKVGEIKLPI	765
Db	569	YKMSQTIFYPFHAGDTYLRVNPQFAPVPGKTDALVRVDFDEHGNAYLENNYKVGEIKLPI	628
Qy	767	PKLNGQTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKENQTDKPSILPQFKRNKAQE	826
Db	629	PKLNGQTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKENQTDKPSILPQFKRNKAQE	688
Qy	827	NSKLDKVEPKTSEKVEKEKISSETGNSTNSLTLEEVPTVDVQKVAKFAESYGMKLEN	886
Db	689	NSKLDKVEPKTSEKVEKEKISSETGNSTNSLTLEEVPTVDVQKVAKFAESYGMKLEN	748
Qy	887	VLENMGDTIELYLPSEGVILKKNMADFTGGAPOGNGENKPSSENGKVSTGTGTVENQPTENKPA	946
Db	749	VLENMGDTIELYLPSEGVILKKNMADFTGGAPOGNGENKPSSENGKVSTGTGTVENQPTENKPA	808
Qy	947	DSLPEAPNEKPVKPNSTONGMLNPEGNYGSDPMLDPALEAPAVDPVQEKLEKFTASYG	1006
Db	809	DSLPEAPNEKPVKPNSTONGMLNPEGNYGSDPMLDPALEAPAVDPVQEKLEKFTASYG	868
Qy	1007	LGLDVSIFNMGDTIELRPSGEVIKKNLSDFTA 1039	
Db	869	LGLDVSIFNMGDTIELRPSGEVIKKNLSDFTA 901	
RESULT 30			
AAU84076			
ID	AAU84076 standard; Peptide; 902 AA.		
XX			
AC	AAU84076;		
XX			
DT	08-MAY-2002 (first entry)		
XX			
DE	S. pneumoniae derived chimeric peptide, VP119.		
XX			
KW	BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;		
KW	pneumonia; streptococcal bacterial infection; mutant; muten;		
KW	BVH-11-2.		
XX			
OS	Streptococcus pneumoniae.		
XX			
FN	WC200198334-A2.		
XX			
PD	27-DEC-2001.		
XX			
PF	19-JUN-2001; 2001WO-CA00908.		
XX			
PR	20-JUN-2000; 2000US-212683P.		
XX			
XX	(SHIR-) SHIRE BIOCHEM INC.		
PA			
XX			
PI	Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;		
XX			
DR	WPI; 2002-122272/16.		
XX			
PT	New Streptococcus pneumoniae BVH-3 and BVH-11 variant and		
PT	epitope-bearing polypeptides, useful as vaccine components for treatin		
PT	or preventing streptococcal infections such as otitis media,		
XX	meningitis, and bacteraemia		
PS	Claim 1; Page ; 113pp; English.		
XX			
CC	The invention describes an isolated polypeptide (I) with 70-90%		
CC	identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of		
CC	BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (I		
CC	comprising (I) is useful for therapeutic or prophylactic treatment of		

CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (i) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX Sequence 902 AA;

Query Match 55.1%; Score 573; DB 23; Length 902;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKDLTEQIAKQKHLVEVKTSHNGLDSSSHEDQYPCGNKEMKDLKKIEKIAGI 526
DB 330 YFFKDLTEQIAKQKHLVEVKTSHNGLDSSSHEDQYPCGNKEMKDLKKIEKIAGI 389
QY 527 MKQYGVKRESIVVKNKNAIIPHGDDHHPIDHKKPVGIGHSHSNYELFKPEGVAKK 586
DB 390 MKQYGVKRESIVVKNKNAIIPHGDDHHPIDHKKPVGIGHSHSNYELFKPEGVAKK 449
QY 587 EGNKVTGTELTNNVNLKNSFNQNTFLANGQKRVSPFPELEKKGINMLVKLITP 646
DB 450 EGNKVTGTELTNNVNLKNSFNQNTFLANGQKRVSPFPELEKKGINMLVKLITP 509
QY 647 DGKLVKSVGKVFEGVGNIAFELDQPLPGQTKFYTIASKDYPEVSYDGTFTVPTSLA 706
DB 510 DGKLVKSVGKVFEGVGNIAFELDQPLPGQTKFYTIASKDYPEVSYDGTFTVPTSLA 569
QY 707 YKASQTIIPYHAGDTYLRVNPQFAVPKGTALYRVDFEFGNAYLENNYKVGSIKLP 766
DB 570 YKASQTIIPYHAGDTYLRVNPQFAVPKGTALYRVDFEFGNAYLENNYKVGSIKLP 629
QY 767 PKLNGQTTTAGNKIPVTMANAYLDNQSTYIVPEVPILEKENQTKPSILPQFKRNKQAE 826
DB 630 PKLNGQTTTAGNKIPVTMANAYLDNQSTYIVPEVPILEKENQTKPSILPQFKRNKQAE 689
QY 827 NSKLDKVEEPTKSEKVEKLESETGNSNSTLEVPVTPVQKVAKFAESYGMKLEN 886
DB 690 NSKLDKVEEPTKSEKVEKLESETGNSNSTLEVPVTPVQKVAKFAESYGMKLEN 749
QY 887 VLFNMDGTIELYLPSEVIVKKNMADFTGAPOGNGENKPSNGKYSTGTVENQPTENKPA 946
DB 750 VLFNMDGTIELYLPSEVIVKKNMADFTGAPOGNGENKPSNGKYSTGTVENQPTENKPA 809
QY 947 DSLPEAPNEKPKVPKNSNTONGMLNPGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
DB 810 DSLPEAPNEKPKVPKNSNTONGMLNPGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYG 869
QY 1007 LGLDSVIFNMDGTIELYLPSEVIVKKNLSDFTA 1039
DB 870 LGLDSVIFNMDGTIELYLPSEVIVKKNLSDFTA 902

RESULT 31

AAU84078

ID AAU84078 standard; Peptide; 902 AA.

XX AAU84078;

AC AAU84078;

XX 08-MAY-2002 (first entry)

XX S. pneumoniae derived chimeric peptide, VP121.
DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX Streptococcus pneumoniae.
OS Synthetic.
OS WO200198334-A2.
PN 27-DEC-2001.
PD 19-JUN-2001; 2001WO-CA00908.
PF 20-JUN-2000; 2000US-212683P.
PR (SHIR-) SHIRE BIOCHEM INC.
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
PI WPI; 2002-122272/16.
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia.
PS Claim 1; Page : 113pp; English.
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumoniae nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX Sequence 902 AA;

Query Match 55.1%; Score 573; DB 23; Length 902;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKDLTEQIAKQKHLVEVKTSHNGLDSSSHEDQYPCGNKEMKDLKKIEKIAGI 526
DB 330 YFFKDLTEQIAKQKHLVEVKTSHNGLDSSSHEDQYPCGNKEMKDLKKIEKIAGI 389
QY 527 MKQYGVKRESIVVKNKNAIIPHGDDHHPIDHKKPVGIGHSHSNYELFKPEGVAKK 586
DB 390 MKQYGVKRESIVVKNKNAIIPHGDDHHPIDHKKPVGIGHSHSNYELFKPEGVAKK 449
QY 587 EGNKVTGTELTNNVNLKNSFNQNTFLANGQKRVSPFPELEKKGINMLVKLITP 646
DB 450 EGNKVTGTELTNNVNLKNSFNQNTFLANGQKRVSPFPELEKKGINMLVKLITP 509
QY 647 DGKLVKSVGKVFEGVGNIAFELDQPLPGQTKFYTIASKDYPEVSYDGTFTVPTSLA 706

Db 510 DGVLEKVSQVFGEGVGNIAFELDPQYLPQGTQFYTIASKDYPEVSYDGTFTVPTSLA 569
QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNAYLENNYKVGKIKLPI 766
Db 570 YKMASQTIFFPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNAYLENNYKVGKIKLPI 629
QY 767 PKLNQGTTRTAGNKIPVTMANAYLDNQSTYIIVEVPILKKNQTKPSILPOFKRNKAQE 826
Db 630 PKLNQGTTRTAGNKIPVTMANAYLDNQSTYIIVEVPILKKNQTKPSILPOFKRNKAQE 689
QY 827 NSKLDKVEEPTSEKVEKEKLSGTSNSTLEEVPTVQVQKVAFAESYGNKLEN 886
Db 690 NKLDEKVEEPTSEKVEKEKLSGTSNSTLEEVPTVQVQKVAFAESYGNKLEN 749
QY 887 VLFNMDGTIELRPSGEVIKKNMADFTGEAPOGNGENKPSGKSTGTGVNQPTENKPA 946
Db 750 VLFNMDGTIELRPSGEVIKKNMADFTGEAPOGNGENKPSGKSTGTGVNQPTENKPA 809
QY 947 DSLPEAPNEKPKPENSTNGMLNPGNGVSDPMLDPALEAPVDPVQEKLEKFTASYG 1006
Db 810 DSLPEAPNEKPKPENSTNGMLNPGNGVSDPMLDPALEAPVDPVQEKLEKFTASYG 869
QY 1007 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
Db 870 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 902
RESULT 32
AAB12731
ID AAB12731 standard; Protein; 568 AA.
XX AC AAB12731;
XX DT 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae NEW1 protein antigen SEQ ID NO:64.
XX KW Streptococcus pneumoniae: BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KW Proprietary; therapy; infection; diagnosis; meningitis; bacteraemia;
XX KW Otitis media; pneumonia; immunisation; bactericidal.
XX OS Streptococcus pneumoniae.
XX PN WO200039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA01218.
XX PR 23-DEC-1998; 98US-0113800.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX DR WPI; 2000-452397/39.
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT Otitis media, bacteraemia and/or pneumonia -
XX PS Claim 18; Fig 29; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens
XX CC have bactericidal activity. The nucleic acids, encoding the protein
XX CC antigens, may be used for the recombinant production of the proteins
XX CC they encode. The protein antigens may then be used as vaccines for the
XX CC prevention and treatment of Streptococcal infections in mammals
XX CC (especially humans) which result in, e.g. meningitis, otitis media,
XX CC bacteraemia and/or pneumonia. The present sequence represents the
XX CC S. pneumoniae NEW1 protein antigen.

SQ Sequence 568 AA;
Query Match 54.7%; Score 568; DB 21; Length 568;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DLTEEQIKAAQKHLEEVKTSNGLSLSSHEQDYPGNAKEMKDLCKIEEKIAGIMQYQ 531
Db 1 DLTEEQIKAAQKHLEEVKTSNGLSLSSHEQDYPGNAKEMKDLCKIEEKIAGIMQYQ 60
QY 532 VKRESIVVKNKNAIIPHGDDHHADPIDHKPVGIGHSHSNHYELFKPEEGVAKKEGKV 591
Db 61 VKRESIVVKNKNAIIPHGDDHHADPIDHKPVGIGHSHSNHYELFKPEEGVAKKEGKV 120
QY 592 YTGELTNVWLLKNSFTNNQNTFLANGQKRVSEFFPELEKLGINMLVKLIIPDGKVL 651
Db 121 YTGELTNVWLLKNSFTNNQNTFLANGQKRVSEFFPELEKLGINMLVKLIIPDGKVL 180
QY 652 EKVSQKVEGEGVGNIAFELDPQYLPQGTQFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 711
Db 181 EKVSQKVEGEGVGNIAFELDPQYLPQGTQFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 240
QY 712 QTIFYPPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNAYLENNYKVGKIKLPIKLNQ 771
Db 241 QTIFYPPHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNAYLENNYKVGKIKLPIKLNQ 300
QY 772 GTTRTAGNKIPVTMANAYLDNQSTYIIVEVPILKKNQTKPSILPOFKRNKAENSKLD 831
Db 301 GTTRTAGNKIPVTMANAYLDNQSTYIIVEVPILKKNQTKPSILPOFKRNKAENSKLD 360
QY 832 EKVEEPTSEKVEKEKLSGTSNSTLEEVPTVQVQKVAFAESYGNKLENVLFNM 891
Db 361 EKVEEPTSEKVEKEKLSGTSNSTLEEVPTVQVQKVAFAESYGNKLENVLFNM 420
QY 892 DGTIELYLPSPGEVIKKNMADFTGEAPOGNGENKPSGKSTGTGVNQPTENKPADSLPE 951
Db 421 DGTIELYLPSPGEVIKKNMADFTGEAPOGNGENKPSGKSTGTGVNQPTENKPADSLPE 480
QY 952 APNEKPKPENSTNGMLNPGNGVSDPMLDPALEAPVDPVQEKLEKFTASYGLGLDS 1011
Db 481 APNEKPKPENSTNGMLNPGNGVSDPMLDPALEAPVDPVQEKLEKFTASYGLGLDS 540
QY 1012 VIFNMDGTIELRPSGEVIKKNLSDFIA 1039
Db 541 VIFNMDGTIELRPSGEVIKKNLSDFIA 568
RESULT 33
AAU83999
ID AAU83999 standard; Peptide; 568 AA.
XX AC AAU83999;
XX DT 08-MAY-2002 (first entry)
XX DE Truncated variant of S. pneumoniae BVH-3, NEW1.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX KW pneumonia; streptococcal bacterial infection; mutant; mutain.
XX OS Streptococcus pneumoniae.
XX OS Synthetic.
XX PN WO200198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.

```
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI: 2002-122272/16.
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
XX
XX Claim 1: Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
XX Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or
XX Streptococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a truncate
XX of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
XX described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX
XX Sequence 568 AA:
XX
XX Query Match 54.7%; Score 568; DB 23; Length 568;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 472 DLPEEQTKAAQKHLVEVKTSHNGSLDLSLSEHQDYPGNAXEMKDLDDKKIEKIAGIMKQYV 531
XX DB 1 DLPEEQTKAAQKHLVEVKTSHNGSLDLSLSEHQDYPGNAXEMKDLDDKKIEKIAGIMKQYV 60
XX
XX QY 532 VKRESIVNKEKNAIIPGHGDHHDADIDHKPVGIGHSNHYELFKPEGVAKKEGKV 591
XX DB 61 VKRESIVNKEKNAIIPGHGDHHDADIDHKPVGIGHSNHYELFKPEGVAKKEGKV 120
XX
XX QY 592 YTGELTNVNLKSTFNQNTLANGKRVSFSPPELEKLGITNMLVKLITPDGKYL 651
XX DB 121 YTGELTNVNLKSTFNQNTLANGKRVSFSPPELEKLGITNMLVKLITPDGKYL 180
XX
XX QY 652 EKVSQKVGEGVGNANFELDQPLPGQTKFTIASKDYPVSVYDGTFTVPISLAYKMAS 711
XX DB 181 EKVSQKVGEGVGNANFELDQPLPGQTKFTIASKDYPVSVYDGTFTVPISLAYKMAS 240
XX
XX QY 712 QTIYFPFHAGDTLVNPNQFAVPGKGTDALVRVDFEFGHGNAYLNNYKVGIEKLPKLNQ 771
XX DB 241 QTIYFPFHAGDTLVNPNQFAVPGKGTDALVRVDFEFGHGNAYLNNYKVGIEKLPKLNQ 300
XX
XX QY 772 GTTTRAGNKIPVTMANAYLDNSTIVVEVPILEKENQTDKPSILPQFRNKAQENSKLD 831
XX DB 301 GTTTRAGNKIPVTMANAYLDNSTIVVEVPILEKENQTDKPSILPQFRNKAQENSKLD 360
XX
XX QY 832 EKVEEPTSKVEKEKLSSETGNTSNTSEVPTDVPQVEKAKFAESYGMKLENVLFNM 891
XX DB 361 EKVEEPTSKVEKEKLSSETGNTSNTSEVPTDVPQVEKAKFAESYGMKLENVLFNM 420
XX
XX QY 892 DGTIELYLPSPGEVKKNNMADFTGEAPQNGENKPSNGKSVTGTVENQPTENKPADSLPE 951
XX DB 421 DGTIELYLPSPGEVKKNNMADFTGEAPQNGENKPSNGKSVTGTVENQPTENKPADSLPE 480
XX
XX QY 952 APNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
XX
XX Db 481 APNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 540
XX
XX QY 1012 VIFNMDGTIELRLPSPGEVKKNLSDFFIA 1039
XX DB 541 VIFNMDGTIELRLPSPGEVKKNLSDFFIA 568
XX
XX RESULT 34
XX AAU84059
XX ID AAU84059 standard; Peptide: 913 AA.
XX
XX AC AAU84059;
XX
XX DT 08-MAY-2002 (first entry)
XX
XX DE S. pneumoniae derived chimeric peptide, NEW32.
XX
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumonia; streptococcal bacterial infection; mutant; mutain;
XX BVH-11-2.
XX
XX OS Streptococcus pneumoniae.
XX OS Synthetic.
XX
XX PN W0200198334-A2.
XX
XX PD 27-DEC-2001.
XX
XX PF 19-JUN-2001; 2001WO-CA00908.
XX
XX PR 20-JUN-2000; 2000US-212683P.
XX
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI: 2002-122272/16.
XX
XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
XX
XX Example 1, page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
XX Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or
XX Streptococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a chimeric
XX gene created from fragments and variant fragments of Streptococcus
XX pneumoniae genes, described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX
XX Sequence 913 AA;
XX
XX Query Match 54.7%; Score 568; DB 23; Length 913;
XX Best Local Similarity 100.0%; Pred. No. 0;
```

```
Matches 568: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DLTEQIKAAQHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKTEETAGIMKQYG 531
DB 346 DLTEQIKAAQHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKTEETAGIMKQYG 405
QY 532 VRESIVNKEKNATIIYPHGDHHDPTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKV 591
DB 406 VRESIVNKEKNATIIYPHGDHHDPTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKV 465
QY 592 YTGEEELTNVNLKKNSTFNQNTLANGQKRVSFSPFPELEKLGINMLVLIITPDGKVL 651
DB 466 YTGEEELTNVNLKKNSTFNQNTLANGQKRVSFSPFPELEKLGINMLVLIITPDGKVL 525
QY 652 EKVSQKVFEGVGNANTANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKNAS 711
DB 526 EKVSQKVFEGVGNANTANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKNAS 585
QY 712 QTIFFPFFHAGDTYLRVNPQFAVPKGTDALVRVDFDFHGNAYLENNYKVGEIKLPIPKLNQ 771
DB 586 QTIFFPFFHAGDTYLRVNPQFAVPKGTDALVRVDFDFHGNAYLENNYKVGEIKLPIPKLNQ 645
QY 772 GTTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTDKPSIILPQKRNKAQENSKLD 831
DB 646 GTTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTDKPSIILPQKRNKAQENSKLD 705
QY 832 EKVEEPTSEKVEKEKLSGTSTNSTLEEYPTVDPVQEKVAKFAESYGMKLENVLFNM 891
DB 706 EKVEEPTSEKVEKEKLSGTSTNSTLEEYPTVDPVQEKVAKFAESYGMKLENVLFNM 765
QY 892 DGTIELYLPSEGEVYKKNMADFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPE 951
DB 766 DGTIELYLPSEGEVYKKNMADFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPE 825
QY 952 APNEKPVKPESTNDGMLNPEGVNGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
DB 826 APNEKPVKPESTNDGMLNPEGVNGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 885
QY 1012 VIFNMDGTIELRLPSGEVYKKNLSDFIA 1039
DB 886 VIFNMDGTIELRLPSGEVYKKNLSDFIA 913
RESULT 35
AAU84051
ID AAU84051 standard; Peptide: 999 AA.
AC AAU84051;
XX
DT 08-MAY-2002 (first entry)
XX
DE S. pneumoniae derived chimeric peptide, NEW17.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein;
KW BVH-11-2.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO2001198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
DR
```

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
XX Example 1; Page -; 113pp; English.
PS
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
CC Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX Sequence 999 AA;
SQ

Query Match 54.7%; Score 568; DB 23; Length 999;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DLTEQIKAAQHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKTEETAGIMKQYG 531
DB 432 DLTEQIKAAQHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKTEETAGIMKQYG 491
QY 532 VKRESTIVNKEKNATIIYPHGDHHDPTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKV 591
DB 492 VRESIVNKEKNATIIYPHGDHHDPTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKV 551
QY 592 YTGEEELTNVNLKKNSTFNQNTLANGQKRVSFSPFPELEKLGINMLVLIITPDGKVL 651
DB 552 YTGEEELTNVNLKKNSTFNQNTLANGQKRVSFSPFPELEKLGINMLVLIITPDGKVL 611
QY 652 EKVSQKVFEGVGNANTANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKNAS 711
DB 612 EKVSQKVFEGVGNANTANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKNAS 671
QY 712 QTIFFPFFHAGDTYLRVNPQFAVPKGTDALVRVDFDFHGNAYLENNYKVGEIKLPIPKLNQ 771
DB 672 QTIFFPFFHAGDTYLRVNPQFAVPKGTDALVRVDFDFHGNAYLENNYKVGEIKLPIPKLNQ 731
QY 772 GTTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTDKPSIILPQKRNKAQENSKLD 831
DB 732 GTTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTDKPSIILPQKRNKAQENSKLD 791
QY 832 EKVEEPTSEKVEKEKLSGTSTNSTLEEYPTVDPVQEKVAKFAESYGMKLENVLFNM 891
DB 792 EKVEEPTSEKVEKEKLSGTSTNSTLEEYPTVDPVQEKVAKFAESYGMKLENVLFNM 851
QY 892 DGTIELYLPSEGEVYKKNMADFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPE 951
DB 852 DGTIELYLPSEGEVYKKNMADFTGEAPQNGENKPSGKSTGTVENQPTENKPADSLPE 911
QY 952 APNEKPVKPESTNDGMLNPEGVNGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
DB 912 APNEKPVKPESTNDGMLNPEGVNGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 971

PS	Claim 18; Fig 23; 106pp; English.	
XX	The present invention describes nucleic acids (I) encoding protein	
CC	antigens (II) from Streptococcus pneumoniae. The protein antigens	
CC	have bactericidal activity. The nucleic acids, encoding the protein	
CC	antigens, may be used for the recombinant production of the proteins	
CC	they encode. The protein antigens may then be used as vaccines for the	
CC	prevention and treatment of Streptococcal infections in mammals	
CC	(especially humans) which result in, e.g. meningitis, otitis media,	
CC	bacteraemia and/or pneumonia. The present sequence represents the	
CC	S. pneumoniae NEW12 protein antigen.	
XX		
SQ	Sequence 1057 AA;	
	Query Match 54.7%; Score 568; DB 21; Length 1057;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	472 DITEEQIKAAQKHLEEVKTSNGLDSSSHQDYPGNAKEMKDLDKKIEEKIAGIMKQYG 531	
DB	1 DITEEQIKAAQKHLEEVKTSNGLDSSSHQDYPGNAKEMKDLDKKIEEKIAGIMKQYG 60	
QY	532 VKRESIVVNKKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKV 591	
DB	61 VKRESIVVNKKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKV 120	
QY	592 YNGEELTNVNVLLKSTFNQNTLANGOKRVSPFPEPELEKKGINMLVLIPTDQKVL 651	
DB	121 YNGEELTNVNVLLKSTFNQNTLANGOKRVSPFPEPELEKKGINMLVLIPTDQKVL 180	
QY	652 ERVSGKVFEGVGNIANFELDQPYLPQGTFTFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 711	
DB	181 ERVSGKVFEGVGNIANFELDQPYLPQGTFTFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 240	
QY	712 QTIFFPFFHAGDTYLRVNPQFAVPKGTDLRVDFEFHGNAYLENNYKVEIKLPIPKLNQ 771	
DB	241 QTIFFPFFHAGDTYLRVNPQFAVPKGTDLRVDFEFHGNAYLENNYKVEIKLPIPKLNQ 300	
QY	772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRNAQENSKLD 831	
DB	301 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRNAQENSKLD 360	
QY	832 EKVEPKTSEKVEKELSTGNTSTNSILEEVPVTPVQEVAKFAESYGMKLENVLFNM 891	
DB	361 EKVEPKTSEKVEKELSTGNTSTNSILEEVPVTPVQEVAKFAESYGMKLENVLFNM 420	
QY	892 DGTIELYLPSEGEVVKKNMADFTGEAPQNGENKPSNGKVGSTGTVENOPTENKPADSLPE 951	
DB	421 DGTIELYLPSEGEVVKKNMADFTGEAPQNGENKPSNGKVGSTGTVENOPTENKPADSLPE 480	
QY	952 APNEKPVKPNSTDNMLNPEGNVSGDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011	
DB	481 APNEKPVKPNSTDNMLNPEGNVSGDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 540	
QY	1012 VIFNMDGTIELPLSPGEVVKKNLSDFIA 1039	
DB	541 VIFNMDGTIELPLSPGEVVKKNLSDFIA 568	
	RESULT 38	
ID	AAU84097	
XX	AAU84097 standard; Peptide; 1058 AA.	
AC	AAU84097;	
DT	08-MAY-2002 (first entry)	
XX	S. pneumoniae derived chimeric peptide, NEW 17.	
XX	BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;	
KW	pneumonia; streptococcal bacterial infection; mutant; mutein;	
KW	BVH-11-2.	
XX		

OS	Streptococcus pneumoniae.	
PN	WO200198334-A2.	
XX		
PD	27-DEC-2001.	
XX		
PF	19-JUN-2001; 2001WO-CA00908.	
XX		
PR	20-JUN-2000; 2000US-212683P.	
XX		
PA	(SHIR-) SHIRE BIOCHEM INC.	
XX		
PI	Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;	
XX		
DR	WPI; 2002-122272/16.	
XX		
PT	New Streptococcus pneumoniae BVH-3 and BVH-11 variant and	
PT	epitope-bearing polypeptides, useful as vaccine components for treating	
PT	or preventing streptococcal infections such as otitis media,	
PT	meningitis, and bacteraemia	
XX		
PS	Example 1; Page -: 113pp; English.	
XX		
CC	The invention describes an isolated polypeptide (I) with 70-90%	
CC	identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of	
CC	BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)	
CC	comprising (I) is useful for therapeutic or prophylactic treatment of	
CC	meningitis, otitis media, bacteraemia or pneumonia infection in an	
CC	individual susceptible to these disorders. (II) is also useful for	
CC	therapeutic or prophylactic treatment of any streptococcal bacterial	
CC	infection (e.g., caused by Streptococcus pneumoniae, group A	
CC	Streptococcus such as Streptococcus pyogenes, group B Streptococcus such	
CC	as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or	
CC	Staphylococcus aureus) in an individual susceptible to the infection.	
CC	A polynucleotide (III) encoding (I) is useful in DNA immunisation	
CC	techniques. The Streptococcus polypeptides are useful in a diagnostic	
CC	test for S. pneumoniae infection. (III) is useful for designing DNA	
CC	probes for use in detecting the presence of Streptococcus in a biological	
CC	sample suspected of containing the bacteria. The DNA probes may also be	
CC	used for detecting circulating S. pneumonia nucleic acid in a sample for	
CC	diagnosing streptococcal infections. This sequence represents a chimeric	
CC	gene created from fragments and variant fragments of Streptococcus	
CC	pneumoniae genes, described in the method of the invention.	
CC	Note: This sequence does not appear in the specification but has	
CC	been created according to information given in the invention.	
XX		
SQ	Sequence 1058 AA;	
	Query Match 54.7%; Score 568; DB 23; Length 1058;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	472 DITEEQIKAAQKHLEEVKTSNGLDSSSHQDYPGNAKEMKDLDKKIEEKIAGIMKQYG 531	
DB	2 DITEEQIKAAQKHLEEVKTSNGLDSSSHQDYPGNAKEMKDLDKKIEEKIAGIMKQYG 61	
QY	532 VKRESIVVNKKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKV 591	
DB	62 VKRESIVVNKKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKV 121	
QY	592 YTGELTNVNVLLKSTFNQNTLANGOKRVSPFPEPELEKKGINMLVLIPTDQKVL 651	
DB	122 YTGELTNVNVLLKSTFNQNTLANGOKRVSPFPEPELEKKGINMLVLIPTDQKVL 181	
QY	652 EKVSKEVGEVGNIANFELDQPYLPQGTFTFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 711	
DB	182 EKVSKEVGEVGNIANFELDQPYLPQGTFTFYTIASKDYPEVSYDGTFTVPTSLAYKMAS 241	
QY	712 QTIFFPFFHAGDTYLRVNPQFAVPKGTDLRVDFEFHGNAYLENNYKVEIKLPIPKLNQ 771	
DB	242 QTIFFPFFHAGDTYLRVNPQFAVPKGTDLRVDFEFHGNAYLENNYKVEIKLPIPKLNQ 301	
QY	772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRNAQENSKLD 831	

|||||
Db 302 GTTRAGNKIPVTFMANAYLDNQSYIIVEPILEREKNOTDKPSILPQFKRKAQNSKLD 361
QY 832 EKVEPKTSEKVEKEKLSGTGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 891
Db 362 EKVEPKTSEKVEKEKLSGTGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 421
QY 892 DGTIELYLPDSGEVVKNNADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPE 951
Db 422 DGTIELYLPDSGEVVKNNADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPE 481
QY 952 APNEKPKPENSTNGMLNPEGNVSDPMLDPALEEAFAVDPVQEKLEKFTASYGLGLDS 1011
Db 482 APNEKPKPENSTNGMLNPEGNVSDPMLDPALEEAFAVDPVQEKLEKFTASYGLGLDS 541
QY 1012 VIFNMDGTIELRLPSGEVVKNNLSDFIA 1039
Db 542 VIFNMDGTIELRLPSGEVVKNNLSDFIA 569
RESULT 39
AAU84058
ID AAU84058 standard; Peptide; 1126 AA.
AC AAU84058;
XX
XX
XX
DT 08-MAY-2002 (first entry)
DE S. pneumoniae derived chimeric peptide, NEW31.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutain;
KW BVH-11-2.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
PN WO200198334-A2.
27-DEC-2001.
19-JUN-2001; 2001WO-CA00908.
20-JUN-2000; 2000US-212683P.
(SHIR-) SHIRE BIOCHEM INC.
Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
WPI; 2002-122272/16.
New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
epitope-bearing polypeptides, useful as vaccine components for treating
or preventing streptococcal infections such as otitis media,
meningitis, and bacteraemia
Example 1; Page -: 113pp; English.
The invention describes an isolated polypeptide (I) with 70-90%
identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
comprising (I) is useful for therapeutic or prophylactic treatment of
meningitis, otitis media, bacteraemia or pneumonia infection in an
individual susceptible to these disorders. (II) is also useful for
therapeutic or prophylactic treatment of any streptococcal bacterial
infection (e.g., caused by Streptococcus pneumoniae, group A
Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
Staphylococcus aureus) in an individual susceptible to the infection.
A polynucleotide (III) encoding (I) is useful in DNA immunisation
techniques. The Streptococcus polypeptides are useful in a diagnostic
test for S. pneumoniae infection. (III) is useful for designing DNA
probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be
used for detecting circulating S. pneumonia nucleic acid in a sample for
diagnosing streptococcal infections. This sequence represents a chimeric
gene created from fragments and variant fragments of Streptococcus
pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 1126 AA;
Query Match 54.7%; Score 568; DB 23; Length 1126;
Best local similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDRKIEKIAIGIMKQYQ 531
Db 559 DLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDRKIEKIAIGIMKQYQ 618
QY 532 VKRESIVVNKEKNAIYPHGDHHDADIDEHKPVGIGHSESNTELPKPEGVAKKGNKV 591
Db 619 VKRESIVVNKEKNAIYPHGDHHDADIDEHKPVGIGHSESNTELPKPEGVAKKGNKV 678
QY 592 YTGELTNVNNLLKNSTFNQNTFLANGQKRVSPFPPELEKLGINMLVKLITPDGKVL 651
Db 679 YTGELTNVNNLLKNSTFNQNTFLANGQKRVSPFPPELEKLGINMLVKLITPDGKVL 738
QY 652 EKSGKVFEGCGVGNIAFELDPQYLPQOTFKYTIASKDYPEVSDGTFTVPTSLAYKMAS 711
Db 739 EKSGKVFEGCGVGNIAFELDPQYLPQOTFKYTIASKDYPEVSDGTFTVPTSLAYKMAS 798
QY 712 QTIYFPFHAGDTYLRVNPQFAVPGKTDALYRVDEFPHGNAYLENNKVGKIKLPKLNQ 771
Db 799 QTIYFPFHAGDTYLRVNPQFAVPGKTDALYRVDEFPHGNAYLENNKVGKIKLPKLNQ 858
QY 772 GTTRAGNKIPVTFMANAYLDNQSYIIVEPILEREKNOTDKPSILPQFKRKAQNSKLD 831
Db 859 GTTRAGNKIPVTFMANAYLDNQSYIIVEPILEREKNOTDKPSILPQFKRKAQNSKLD 918
QY 832 EKVEPKTSEKVEKEKLSGTGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 891
Db 919 EKVEPKTSEKVEKEKLSGTGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 978
QY 892 DGTIELYLPDSGEVVKNNADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPE 951
Db 979 DGTIELYLPDSGEVVKNNADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPE 1038
QY 952 APNEKPKPENSTNGMLNPEGNVSDPMLDPALEEAFAVDPVQEKLEKFTASYGLGLDS 1011
Db 1039 APNEKPKPENSTNGMLNPEGNVSDPMLDPALEEAFAVDPVQEKLEKFTASYGLGLDS 1098
QY 1012 VIFNMDGTIELRLPSGEVVKNNLSDFIA 1039
Db 1099 VIFNMDGTIELRLPSGEVVKNNLSDFIA 1126
RESULT 40
AAU84055
ID AAU84055 standard; Peptide; 1139 AA.
AC AAU84055;
XX
XX
DT 08-MAY-2002 (first entry)
DE S. pneumoniae derived chimeric peptide, NEW28.
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutain;
KW BVH-11-2.
XX Streptococcus pneumoniae.
OS Synthetic.
PN WO200198334-A2.

XX 27-DEC-2001.
PD
XX PF
XX
PR 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
DR
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Example 1; Page : 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
CC Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ Sequence 1139 AA:
Query Match 54.7%; Score 568; DB 23; Length 1139;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 472 DLTEEQIKAAQKHLEEVKTSNGLDSSSHQDYPGNAKEMKDLKRIEETIAGIMKQYG 531
DB 572 DLTEEQIKAAQKHLEEVKTSNGLDSSSHQDYPGNAKEMKDLKRIEETIAGIMKQYG 631
QY 532 VKRESIVVNEKNATIIYPGHHADPTIDEHKPVGIGHSHSNYELFKPEEGVAKKGNV 591
DB 632 VKRESIVVNEKNATIIYPGHHADPTIDEHKPVGIGHSHSNYELFKPEEGVAKKGNV 691
QY 592 YTGELTNVNLKSTNNFTLANGQKRVSFSPPELEKLGINMLVKLIIPDGKVL 651
DB 692 YTGELTNVNLKSTNNFTLANGQKRVSFSPPELEKLGINMLVKLIIPDGKVL 751
QY 652 EKVSQKVEGNGNIANFELQDYPILPGQTFYITASKDYPEVSDGTFVTPTSILAYKMAS 711
DB 752 EKVSQKVEGNGNIANFELQDYPILPGQTFYITASKDYPEVSDGTFVTPTSILAYKMAS 811
QY 712 QTIFPFFHAGTYLRVNPQFAPVPGTDLVRVDFEFGHNAVLENNYKVGKELPIPKLQ 771
DB 812 QTIFPFFHAGTYLRVNPQFAPVPGTDLVRVDFEFGHNAVLENNYKVGKELPIPKLQ 871
QY 772 GTTRTAGNKIPVTFEMANAYLDNQSTYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLD 831
DB 872 GTTRTAGNKIPVTFEMANAYLDNQSTYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLD 931

QY 832 EKVEEPTSEKVEKELSETGNSNSTLEEPTVDVQEKVAKFAESYGMKLENVLFNN 891
DB 932 EKVEEPTSEKVEKELSETGNSNSTLEEPTVDVQEKVAKFAESYGMKLENVLFNN 991
QY 892 DGTIELYLPSPGEVIKKNMADFTGEAPGNGENKPSNGKYSTGTVENQPTENKPADSLPE 951
DB 992 DGTIELYLPSPGEVIKKNMADFTGEAPGNGENKPSNGKYSTGTVENQPTENKPADSLPE 1051
QY 952 APNEKPVKPNSTDNGLNPNPEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
DB 1052 APNEKPVKPNSTDNGLNPNPEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1111
QY 1012 VIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
DB 1112 VIFNMDGTIELRLPSGEVIKKNLSDFIA 1139
RESULT 41
AA084006
ID AA084006 standard; Peptide; 632 AA.
XX
AC AA084006;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-3, NEW53.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutain.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Claim 1; Page : 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX
SQ Sequence 632 AA;

Query Match 51.1%; Score 531; DB 23; Length 632;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 631; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 408 QICQPTLPNNSLATPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSGHDHNY 467
DB 1 QICQPTLPNNSLATPSLPINPGTSHKHEEDGYGFDANRIIAEDSGFVMSGHDHNY 60
QY 468 FFKKOLTEQIKAAQKHLEEVKTSINGLSDSSSHEDQYPGNAKEMKLDKKIEKIKAGIM 527
DB 61 FFKKOLTEQIKAAQKHLEEVKTSINGLSDSSSHEDQYPGNAKEMKLDKKIEKIKAGIM 120
QY 528 KOYGKRESIVVNEKNALIIYPHGDHHDPTDEHKPVGIGSHSNYELFKPEEGVAKKE 587
DB 121 KOYGKRESIVVNEKNALIIYPHGDHHDPTDEHKPVGIGSHSNYELFKPEEGVAKKE 180
QY 588 GKNVYTGELTNVWLLKNSTFNQNFNLANGQKRVSFSPPELEKLGIMLVKLTIPD 647
DB 181 GKNVYTGELTNVWLLKNSTFNQNFNLANGQKRVSFSPPELEKLGIMLVKLTIPD 240
QY 648 GKVLKVSCKVFGGEGVGNIANFELDQYLPQGTFKYTIASKDYPEVSDGTFTVPTSLAY 707
DB 241 GKVLKVSCKVFGGEGVGNIANFELDQYLPQGTFKYTIASKDYPEVSDGTFTVPTSLAY 300
QY 708 KMASQTIFFPFHAGDTYLRVNPQFVAPVPGTALVRVDFEFGHGNAYLENNYKVGKILPIP 767
DB 301 KMASQTIFFPFHAGDTYLRVNPQFVAPVPGTALVRVDFEFGHGNAYLENNYKVGKILPIP 360
QY 768 KLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKEKNTQDKPSLPOFKRNKAQEN 827
DB 361 KLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKEKNTQDKPSLPOFKRNKAQEN 420
QY 828 SKLDEKVEEPTSEKVEKEKLSQNSTLEEVPTVDPVQEKVAKFAESYGMKLENV 887
DB 421 SKLDEKVEEPTSEKVEKEKLSQNSTLEEVPTVDPVQEKVAKFAESYGMKLENV 480
QY 888 LFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPSNGKYSTGTGVNQPTENKPAD 947
DB 481 LFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPSNGKYSTGTGVNQPTENKPAD 540
QY 948 SLPEAPNEKPVKPNSTNGMLNPGNVSQDPMPLDPALEAPAVDPVQEKLEKFTASYGL 1007
DB 541 SLPEAPNEKPVKPNSTNGMLNPGNVSQDPMPLDPALEAPAVDPVQEKLEKFTASYGL 600
QY 1008 GLDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 1039
DB 601 GLDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 632

RESULT 42

AAB12719

ID AAB12719 standard; Protein; 528 AA.

XX
AC AAB12719;

XX
DT 21-NOV-2000 (first entry)

XX
DE Streptococcus pneumoniae BVH-3B protein antigen SEQ ID NO:10.

XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.

XX
OS Streptococcus pneumoniae.

XX
PN WO200039299-A2.

XX

PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
WIPI; 2000-452397/39.
DR N-PSDB: AAA65734.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Claim 18; Fig 10; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3B protein antigen.
XX
SQ Sequence 528 AA;

Query Match 50.8%; Score 528; DB 21; Length 528;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 MKDLKKIEEKIAGIKKQYGVKRESIVVNEKNALIIYPHGDHHDPTDEHKPVGIGSH 571
DB 1 MKDLKKIEEKIAGIKKQYGVKRESIVVNEKNALIIYPHGDHHDPTDEHKPVGIGSH 60
QY 572 SNLEFKEPEGVAKKQYGVKRESIVVNEKNALIIYPHGDHHDPTDEHKPVGIGSH 631
DB 61 SNLEFKEPEGVAKKQYGVKRESIVVNEKNALIIYPHGDHHDPTDEHKPVGIGSH 120
QY 632 EKKLGINMLVKLTIPDGKYLEKVSQVFGGEGVGNIANFELDQYLPQGTFKYTIASKDYP 691
DB 121 EKKLGINMLVKLTIPDGKYLEKVSQVFGGEGVGNIANFELDQYLPQGTFKYTIASKDYP 180
QY 692 EVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFVAPVPGTALVRVDFEFGHNA 751
DB 181 EVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFVAPVPGTALVRVDFEFGHNA 240
QY 752 YLENNYKVGKILPIPKNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKEKNTQDK 811
DB 241 YLENNYKVGKILPIPKNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKEKNTQDK 300
QY 812 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSQNSTLEEVPTVDPVQEK 871
DB 301 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSQNSTLEEVPTVDPVQEK 360
QY 872 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPSNGKV 931
DB 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVYKKNMADFTGEAPQNGENKPSNGKV 420
QY 932 STGTGVNQPTENKPADSLPEAPNEKPVKPNSTNGMLNPGNVSQDPMPLDPALEAPAV 991
DB 421 STGTGVNQPTENKPADSLPEAPNEKPVKPNSTNGMLNPGNVSQDPMPLDPALEAPAV 480
QY 992 DPVQEKLEKFTASYGLSDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 1039
DB 481 DPVQEKLEKFTASYGLSDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 528

RESULT 43
AAB12719

ID AU84024 standard; Peptide; 528 AA.
 XX AU84024;
 AC
 XX
 XX 08-MAY-2002 (first entry)
 DT
 XX
 XX Truncated variant of *S. pneumoniae* BVH-3, BVH-3B.
 DE
 XX
 XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutin.
 KW
 XX
 XX Streptococcus pneumoniae.
 OS Synthetic.
 OS
 XX WO2001198334-A2.
 PN
 XX 27-DEC-2001.
 PD
 XX
 XX 19-JUN-2001; 2001WO-CA00908.
 PF
 XX 20-JUN-2000; 2000US-212683P.
 PR
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 PI WPI; 2002-122272/16.
 DR
 XX
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 XX Example 1; Page -; 113pp; English.
 PS
 XX The invention describes an isolated polypeptide (I) with 70-90%
 XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, *S. dysgalactiae*, *S. uberis*, *S. nodocidia* or
 CC *Staphylococcus aureus*) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for *S. pneumoniae* infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating *S. pneumoniae* nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX
 XX Sequence 528 AA;
 SQ
 Query Match 50.8%; Score 528; DB 23; Length 528;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 512 MKDLKKIEETAGTMMQYGVKRESIVVKNENAIYPHGDHHDADPDEHKPVGIGHSH 571
 DB 1 MKDLKKIEETAGTMMQYGVKRESIVVKNENAIYPHGDHHDADPDEHKPVGIGHSH 60
 QY 572 SNYELFKPEEGVAKREGNKVYTGELTNVLLKNSTNNONFTLANGKQKVSFFPEL 631
 DB 61 SNYELFKPEEGVAKREGNKVYTGELTNVLLKNSTNNONFTLANGKQKVSFFPEL 120
 QY 632 EKKLGINMLVKLITPDGKVLKSGVFGGVGNIANFELDQPLPQCTFKYTIASKDYP 691

DB 121 EKKLGINMLVKLITPDGKVLKSGVFGGVGNIANFELDQPLPQCTFKYTIASKDYP 180
 QY 692 EVSYDGTFTVPTSLAYKMASQTIFYPEHAGDTVLRVNPQFAVPGKTDALVRVDFEHGNA 751
 DB 181 EVSYDGTFTVPTSLAYKMASQTIFYPEHAGDTVLRVNPQFAVPGKTDALVRVDFEHGNA 240
 QY 752 YLENNYKVGELKPIPKLNOGTTTACNKIPVTFMANAYLDNQSTIYVEVPILLEKENQTD 811
 DB 241 YLENNYKVGELKPIPKLNOGTTTACNKIPVTFMANAYLDNQSTIYVEVPILLEKENQTD 300
 QY 812 KPSILPQFKRNKAGENSKLDEKVEEPTSEKVEKEKLESETGNSTNSLTLEEVTVDPVQE 871
 DB 301 KPSILPQFKRNKAGENSKLDEKVEEPTSEKVEKEKLESETGNSTNSLTLEEVTVDPVQE 360
 QY 872 KVAKFAESYGNKLENVLFNMDGTIELYLPAGEVIKKMADFTGEAPQNGENKPSSENGKV 931
 DB 361 KVAKFAESYGNKLENVLFNMDGTIELYLPAGEVIKKMADFTGEAPQNGENKPSSENGKV 420
 QY 932 STGTVENOPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEGNVSGDPMLEDPALAEAPAV 991
 DB 421 STGTVENOPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEGNVSGDPMLEDPALAEAPAV 480
 QY 992 DPVOEKLEKFTASYGLGDSVIFNMDGTIELRLPSSGEVIKKNLSDFIA 1039
 DB 481 DPVOEKLEKFTASYGLGDSVIFNMDGTIELRLPSSGEVIKKNLSDFIA 528
 RESULT 44
 AAB12724
 ID AAB12724 standard; Protein; 509 AA.
 XX
 XX AAB12724;
 AC
 XX 21-NOV-2000 (first entry)
 DT
 XX Streptococcus pneumoniae L-BVH-3-AD protein antigen SEQ ID NO:57.
 DE
 XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW Otitis media; pneumonia; immunisation; bactericidal.
 XX
 XX Streptococcus pneumoniae.
 OS
 XX WO200039299-A2.
 PN
 XX 06-JUL-2000.
 PD
 XX 20-DEC-1999; 99WO-CA01218.
 PF
 XX 23-DEC-1998; 98US-0113800.
 PR
 XX (BIOC-) BIOCHEM PHARMA INC.
 PA
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 PI WPI; 2000-452397/39.
 DR
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 PT
 XX Claim 18; Fig 22; 106pp; English.
 PS
 XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the proteins
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC *S. pneumoniae* L-BVH-3-AD protein antigen.
 CC XX

SQ Sequence 509 AA;
 Query Match 49.0%; Score 509; DB 21; Length 509;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKFSKYYIAAGSAVIVSLGAYALNQRHSGENKDNRRVSYVDGSSOSSKSENLTDPQVS	60
Db	1	MKFSKYYIAAGSAVIVSLGAYALNQRHSGENKDNRRVSYVDGSSOSSKSENLTDPQVS	60
Qy	61	QKEGIOAEQIVKITDOGYVTSHGHDHYHNGKVPYDALFSELLMKDNYQLKDADIYN	120
Db	61	QKEGIOAEQIVKITDOGYVTSHGHDHYHNGKVPYDALFSELLMKDNYQLKDADIYN	120
Qy	121	EVKGGYIIKVDGKYYVYLKDAHAADNVRTKDEINRQKQEHVKDNEKVNNAVAVARSQGRY	180
Db	121	EVKGGYIIKVDGKYYVYLKDAHAADNVRTKDEINRQKQEHVKDNEKVNNAVAVARSQGRY	180
Qy	181	TTNDGVVFNPADIIEDTGNAYIVPHGGHYHYIPKSDLSASELAAGAHLAGKNMOPSQLS	240
Db	181	TTNDGVVFNPADIIEDTGNAYIVPHGGHYHYIPKSDLSASELAAGAHLAGKNMOPSQLS	240
Qy	241	YSSTASDNNQTSVAGKSTSPANKSENQLSKLXELYDPSAQRYSSDGLVFPDAKIISR	300
Db	241	YSSTASDNNQTSVAGKSTSPANKSENQLSKLXELYDPSAQRYSSDGLVFPDAKIISR	300
Qy	301	TPNGVAIPGHGDHYHPIPKSLSALEEKIARMPVLSGTGTVSTNAKPNVSVSLGSLSEN	360
Db	301	TPNGVAIPGHGDHYHPIPKSLSALEEKIARMPVLSGTGTVSTNAKPNVSVSLGSLSEN	360
Qy	361	PSLSTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPNNSLA	420
Db	361	PSLSTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPNNSLA	420
Qy	421	TPSPSPINPGTSHKEKHEEDGYGFDANRIAEDESFGFVMSHGDNHYFFPKDLTEEQIKA	480
Db	421	TPSPSPINPGTSHKEKHEEDGYGFDANRIAEDESFGFVMSHGDNHYFFPKDLTEEQIKA	480
Qy	481	AQKHLREVKTSHNGLDLSLSSHEQDYPGNA 509	
Db	481	AQKHLREVKTSHNGLDLSLSSHEQDYPGNA 509	

RESULT 45
 AAU84023
 ID AAU84023 standard; Peptide; 509 AA.
 AC AAU84023;
 XX
 XX
 DT 08-MAY-2002 (first entry)
 DE Truncated variant of *S. pneumoniae* BVH-3, L-BVH-3AD.
 DE BVH-3; BVH-11; vaccine; meningitis; Otitis media; Bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein.
 KW
 XX Streptococcus pneumoniae.
 OS Synthetic.
 OS
 XX WO2001198334-A2.
 PN
 XX
 PD 27-DEC-2001.
 XX
 XX 19-JUN-2001; 2001WO-CA00909.
 XX
 XX 20-JUN-2000; 2000US-212683P.
 XX
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX
 XX NPI; 2002-122272/16.
 DR
 XX

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia

Example 1; Page -: 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.

Note: This sequence does not appear in the specification but has been created according to information given in the invention.

XX Sequence 509 AA;

Query Match 49.0%; Score 509; dB 23; Length 509;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKFSKKYIAAGSAIVISLSLCAYALNHRSQENKDNRRVSYVDGSSQSKSENLTDPQVS 60
DB 1 MKFSKKYIAAGSAIVISLSLCAYALNHRSQENKDNRRVSYVDGSSQSKSENLTDPQVS 60

QY 61 QKEGIQAEQIVIKTDGGYVTSHGDDHYHYNGKVPYDALFSEELLKMDPNYQLKDADIVN 120
DB 61 QKEGIQAEQIVIKTDGGYVTSHGDDHYHYNGKVPYDALFSEELLKMDPNYQLKDADIVN 120

QY 121 EVKGGYIKYDGGKYVYLYKDAAHADNRVTKDEINRQXQEHVKDKENKNSNVAVARSQGRY 180
DB 121 EVKGGYIKYDGGKYVYLYKDAAHADNRVTKDEINRQXQEHVKDKENKNSNVAVARSQGRY 180

QY 181 TTNDGYVFNPAADITEDTGNAYIVPHGGHYHYIPKSDLSASELAARAKAHLAGKNQPSOLS 240
DB 181 TTNDGYVFNPAADITEDTGNAYIVPHGGHYHYIPKSDLSASELAARAKAHLAGKNQPSOLS 240

QY 241 YSSTASDNNTQSAKGSTSKPANKSENLOSLLKELYDPSAQRYSSESDGLVDFPAKIISR 300
DB 241 YSSTASDNNTQSAKGSTSKPANKSENLOSLLKELYDPSAQRYSSESDGLVDFPAKIISR 300

QY 301 TPNGVAPLPHGDHYHFIYPYKLSALEEPIARMPVTSGTGSTVSTNAKNEVYSSLGSLSSN 360
DB 301 TPNGVAPLPHGDHYHFIYPYKLSALEEPIARMPVTSGTGSTVSTNAKNEVYSSLGSLSSN 360

QY 361 PSSLTTSKEIASSADGYTFNPKDIVEATATAYIVRHGDHFHYIPKSNQIQOGLTPNNSLA 420
DB 361 PSSLTTSKEIASSADGYTFNPKDIVEATATAYIVRHGDHFHYIPKSNQIQOGLTPNNSLA 420

QY 421 TSPSLPINPCTSHKEHEEDGYGFDANRIIAEDSGFVMSHGDHNHYFFRKDLTEEOIKA 480
DB 421 TSPSLPINPCTSHKEHEEDGYGFDANRIIAEDSGFVMSHGDHNHYFFRKDLTEEOIKA 480

QY 481 AQKHLVEKVTSHNGDLSLSSHEQDYPGNA 509
DB 481 AQKHLVEKVTSHNGDLSLSSHEQDYPGNA 509

Search completed: May 13, 2003, 14:00:04
Job time : 86 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:58:38 ; Search time 26 Seconds
(without alignments)
3841.679 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 1039

Sequence: 1 MKFSKKYIAGSAVIVSLSL.....TELRLPGEVTKNLSDFIA 1039

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 8

Total number of hits satisfying chosen parameters: 45

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database :

PIR_73:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	834	80.3	1039	2 H95115	conserved hypothe
2	632	60.8	1039	2 D97985	hypotheical prote
3	32	3.1	802	2 G95136	conserved domain p
4	32	3.1	819	2 B95136	conserved domain p
5	32	3.1	828	2 E98004	hypotheical prote
6	32	3.1	839	2 G95115	conserved hypothe
7	32	3.1	853	2 C97985	hypotheical prote
8	32	3.1	855	2 D98004	histidine Motif-Q
9	16	1.5	822	2 T46758	hypotheical 92.4K
10	9	0.9	45	2 E97985	hypotheical prote
11	9	0.9	182	2 F97985	hypotheical prote
12	8	0.8	92	2 C69541	conserved hypothe
13	8	0.8	96	2 A61291	ferredoxin [2Fe-2S
14	8	0.8	97	1 FEED	ferredoxin [2Fe-2S
15	8	0.8	97	1 FEYB6	ferredoxin [2Fe-2S
16	8	0.8	127	2 S39689	yycD protein - Bac
17	8	0.8	139	1 PER2	ferredoxin [2Fe-2S
18	8	0.8	146	2 A96919	PTS system, manniti
19	8	0.8	154	2 AD3475	protein-Npi-phosph
20	8	0.8	207	2 A69349	conserved hypothe
21	8	0.8	233	1 R5HG1T	ribosomal protein
22	8	0.8	273	1 R5R22	ribosomal protein
23	8	0.8	273	1 R5M22	ribosomal protein
24	8	0.8	275	2 T11810	ribosomal protein
25	8	0.8	275	2 T07361	ribosomal protein
26	8	0.8	276	2 T07531	ribosomal protein
27	8	0.8	277	2 AF2987	ABC transporter, m
28	8	0.8	278	2 B98296	sulfate ABC transp
29	8	0.8	278	2 A13633	hypotheical prote

30 8 0.8 327 2 H59094
31 8 0.8 345.2 2 G83148
32 8 0.8 410 2 AD2087
33 8 0.8 472 2 F81329
34 8 0.8 522 2 B83735
35 8 0.8 573 2 B97331
36 8 0.8 672 2 S61157
37 8 0.8 753 2 S48261
38 8 0.8 763 2 T18280
39 8 0.8 765 2 T40674
40 8 0.8 813 2 G96494
41 8 0.8 840 2 C75313
42 8 0.8 984 2 G86393
43 8 0.8 1148 2 S72635
44 8 0.8 1218 2 T14265
45 8 0.8 1237 2 T45070

ALIGNMENTS

RESULT 1

H95115

Conserved hypotheical protein SPI004 [imported] - Streptococcus pneumoniae (strain TIGR)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: H95115

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris

A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: H95115

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1039 <AUR>

A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:gl4972477; GSPDB:GN00164; TIGR:

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SPI004

Query Match 80.3%; Score 834; DB 2; Length 1039;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1034; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKFSKKYIAGSAVIVSLSLCAVALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVS 60

Db 1 MKFSKKYIAGSAVIVSLSLCAVALNQHRSQENKNNRVSYVDGSSQSKSENLTDPQVS 60

QY 61 QKESIQAEQIVIKITDQGYVTSHGDIHYNGKVPYDALFSELLMKDPNYQLKDADIVN 120

Db 61 QKESIQAEQIVIKITDQGYVTSHGDIHYNGKVPYDALFSELLMKDPNYQLKDADIVN 120

QY 121 EVKGGYIIVKVDGYYVYLKDAHADNVRTKDEINRQKQEHVKNKVNNAVARSQRY 180

Db 121 EVKGGYIIVKVDGYYVYLKDAHADNVRTKDEINRQKQEHVKNKVNNAVARSQRY 180

QY 181 TTNDGYVFNPAIDTGNAYIIPHGHHYIIPKSDLSASELAHAHLAGKNQFSQLS 240

Db 181 TTNDGYVFNPAIDTGNAYIIPHGHHYIIPKSDLSASELAHAHLAGKNQFSQLS 240

QY 241 YSTASDNNTQSAVKGSTSKPANKSENLSLKLKELVDSQAQRYSDGLVFPFPAKIISR 300

Db 241 YSTASDNNTQSAVKGSTSKPANKSENLSLKLKELVDSQAQRYSDGLVFPFPAKIISR 300

QY 301 TPNGVAIPHGDHYHFFIPYSKLSALEEKIARMVPISGTGTSTVSTNAKPNEVWSLGLSSN 360

Db 301 TPNGVAIPHGDHYHFFIPYSKLSALEEKIARMVPISGTGTSTVSTNAKPNEVWSLGLSSN 360

QY 361 PSSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGDHFHYIPKSNQIGQPTLPNNSLA 420

Db 361 PSSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGDHFHYIPKSNQIGQPTLPNNSLA 420

Db 361 PSSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGDFHIFPKSNQIGQPTLPNNLSA 420
 QY 421 TPSPSLPINCTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDNHRYFFKDLTEEQIKA 480
 Db 421 TPSPSLPINCTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDNHRYFFKDLTEEQIKA 480
 QY 481 AOKHLEEVKTSKGLDSSLSEHODYPCNAKEMKDLDDKTEETAGIMKQYGVKRESIVVN 540
 Db 481 AOKHLEEVKTSKGLDSSLSEHODYPCNAKEMKDLDDKTEETAGIMKQYGVKRESIVVN 540
 QY 541 KEKNAIITPHGDRHHDPIDEHKPVGIGHSHSNLYELFKEEGVAKKEGKVYTGELTNV 600
 Db 541 KEKNAIITPHGDRHHDPIDEHKPVGIGHSHSNLYELFKEEGVAKKEGKVYTGELTNV 600
 QY 601 VNLKSTFNQNTLANGOKRVSPFPELEKKGINMLVLIITPDGKVLKESKGVFG 660
 Db 601 VNLKSTFNQNTLANGOKRVSPFPELEKKGINMLVLIITPDGKVLKESKGVFG 660
 QY 661 EGVGNIANFELDQYPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPFA 720
 Db 661 EGVGNIANFELDQYPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPFA 720
 QY 721 GDTYLRVNPQFVAPKGTDALVRVDFEFGHGNAYLNNYKVEIKLPIPKLNOGTTTAGNK 780
 Db 721 GDTYLRVNPQFVAPKGTDALVRVDFEFGHGNAYLNNYKVEIKLPIPKLNOGTTTAGNK 780
 QY 781 IPVTFMANAYLDNQSTYIVVEPILKENDTKPSILPQKRNKAQENSKLDEKVEEPKTS 840
 Db 781 IPVTFMANAYLDNQSTYIVVEPILKENDTKPSILPQKRNKAQENSKLDEKVEEPKTS 840
 QY 841 EKVEKELSTGNTSNTLEEVPTDPOEKVAKFAESYGMKLEWLFNMDGTIELYLP 900
 Db 841 EKVEKELSTGNTSNTLEEVPTDPOEKVAKFAESYGMKLEWLFNMDGTIELYLP 900
 QY 901 SGEVIRKKNMADFTGEAPQNGENKPSNGKSTGTVENOPTENKPADSLPEAPNEKPKVP 960
 Db 901 SGEVIRKKNMADFTGEAPQNGENKPSNGKSTGTVENOPTENKPADSLPEAPNEKPKVP 960
 QY 961 ENSTDNGMLNPEGVNSDPMLEPALEAPVDPVQEKLEKFTASVGLGLDSVIFNMDGTI 1020
 Db 961 ENSTDNGMLNPEGVNSDPMLEPALEAPVDPVQEKLEKFTASVGLGLDSVIFNMDGTI 1020
 QY 1021 ELRLPSGEVIKKNLSD 1036
 Db 1021 ELRLPSGEVIKKNLSD 1036

RESULT 3

Query Match 60.8%; Score 632; DB 2; Length 1039;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1032; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MKFSKKYIAAGSAVIVSLSCAYALNQHRSQENKNNRVSYVDGSSQSKSENITPDQVS 60
 QY 61 QKREGIAEQIVIKITDOGYVTSGHDHYHYNGKPYDADFSEELMKDPNYQLKADATVN 120
 Db 61 QKREGIAEQIVIKITDOGYVTSGHDHYHYNGKPYDADFSEELMKDPNYQLKADATVN 120
 QY 121 EVKGGYIIVKVDGKYVYLKDAAHADNVRTKDEINQKQEHVKDNEKVNNAVARSQRY 180
 Db 121 EVKGGYIIVKVDGKYVYLKDAAHADNVRTKDEINQKQEHVKDNEKVNNAVARSQRY 180
 QY 181 TTNDGVYFNPADITIEDTGNAYIVPHGGHYHYIPKSDLSASSELAAKAHLAGKNQPSOLS 240
 Db 181 TTNDGVYFNPADITIEDTGNAYIVPHGGHYHYIPKSDLSASSELAAKAHLAGKNQPSOLS 240
 QY 241 YSSTASDNNTQSVAKGSTSKPANKSENLOSLLKELYDPSAQRYSESGLVDFDPAKIISR 300
 Db 241 YSSTASDNNTQSVAKGSTSKPANKSENLOSLLKELYDPSAQRYSESGLVDFDPAKIISR 300
 QY 301 TPRGVAIPHGDRHDFIIPYSKLSALEEKIARRVPISGTGSTVSTNAKPNEVSSLSGSSN 360
 Db 301 TPRGVAIPHGDRHDFIIPYSKLSALEEKIARRVPISGTGSTVSTNAKPNEVSSLSGSSN 360
 QY 361 PSSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGDFHIFPKSNQIGQPTLPNNLSA 420
 Db 361 PSSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGDFHIFPKSNQIGQPTLPNNLSA 420
 QY 421 TPSPSLPINCTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDNHRYFFKDLTEEQIKA 480
 Db 421 TPSPSLPINCTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDNHRYFFKDLTEEQIKA 480
 QY 481 AOKHLEEVKTSKGLDSSLSEHODYPCNAKEMKDLDDKTEETAGIMKQYGVKRESIVVN 540
 Db 481 AOKHLEEVKTSKGLDSSLSEHODYPCNAKEMKDLDDKTEETAGIMKQYGVKRESIVVN 540
 QY 541 KEKNAIITPHGDRHHDPIDEHKPVGIGHSHSNLYELFKEEGVAKKEGKVYTGELTNV 600
 Db 541 KEKNAIITPHGDRHHDPIDEHKPVGIGHSHSNLYELFKEEGVAKKEGKVYTGELTNV 600
 QY 601 VNLKSTFNQNTLANGOKRVSPFPELEKKGINMLVLIITPDGKVLKESKGVFG 660
 Db 601 VNLKSTFNQNTLANGOKRVSPFPELEKKGINMLVLIITPDGKVLKESKGVFG 660
 QY 661 EGVGNIANFELDQYPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPFA 720
 Db 661 EGVGNIANFELDQYPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPFA 720
 QY 721 GDTYLRVNPQFVAPKGTDALVRVDFEFGHGNAYLNNYKVEIKLPIPKLNOGTTTAGNK 780
 Db 721 GDTYLRVNPQFVAPKGTDALVRVDFEFGHGNAYLNNYKVEIKLPIPKLNOGTTTAGNK 780
 QY 781 IPVTFMANAYLDNQSTYIVVEPILKENDTKPSILPQKRNKAQENSKLDEKVEEPKTS 840
 Db 781 IPVTFMANAYLDNQSTYIVVEPILKENDTKPSILPQKRNKAQENSKLDEKVEEPKTS 840
 QY 841 EKVEKELSTGNTSNTLEEVPTDPOEKVAKFAESYGMKLEWLFNMDGTIELYLP 900
 Db 841 EKVEKELSTGNTSNTLEEVPTDPOEKVAKFAESYGMKLEWLFNMDGTIELYLP 900
 QY 901 SGEVIRKKNMADFTGEAPQNGENKPSNGKSTGTVENOPTENKPADSLPEAPNEKPKVP 960
 Db 901 SGEVIRKKNMADFTGEAPQNGENKPSNGKSTGTVENOPTENKPADSLPEAPNEKPKVP 960
 QY 961 ENSTDNGMLNPEGVNSDPMLEPALEAPVDPVQEKLEKFTASVGLGLDSVIFNMDGTI 1020
 Db 961 ENSTDNGMLNPEGVNSDPMLEPALEAPVDPVQEKLEKFTASVGLGLDSVIFNMDGTI 1020
 QY 1021 ELRLPSGEVIKKNLSD 1036
 Db 1021 ELRLPSGEVIKKNLSD 1036

C95136
conserved domain protein Sp1175 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95136
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95136
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75284.1; PID:gl4972655; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: Sp1175
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 3.1%; Score 32; DB 2; Length 802;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||
Db 52 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 83
|||||

RESULT 4
B95136
conserved domain protein Sp1174 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95136
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-819 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75283.1; PID:gl4972654; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: Sp1174
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 3.1%; Score 32; DB 2; Length 819;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||
Db 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||

RESULT 5
E98004
hypothetical protein phtA [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: E98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: C97985
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99711.1; PID:gl5458514; GSPDB:GN00174
C:Genetics:
A:Gene: phtA

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E98004
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-828 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:gl5458683; GSPDB:GN00174
C:Genetics:
A:Gene: phtA
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 3.1%; Score 32; DB 2; Length 828;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||
Db 78 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 109
|||||

RESULT 6
G95115
conserved hypothetical protein SPI003 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: G95115
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95115
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75120.1; PID:gl4972476; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI003
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 3.1%; Score 32; DB 2; Length 839;
Best Local Similarity 100.0%; Pred. No. 2.6e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||
Db 67 AEQIVIKITDQGYVTSHGDRHYNGKVPYDA 98
|||||

RESULT 7
C97985
hypothetical protein phtD [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: C97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: C97985
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99711.1; PID:gl5458514; GSPDB:GN00174
C:Genetics:
A:Gene: phtD

```

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein
Query Match          3.1%; Score 32; DB 2; Length 853;
Best Local Similarity 100.0%; Pred. No. 2.6e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 AEQIVIKITDQGYVTSRSHGHHYNGKVPYDA 98
      |||||
Db 67 AEQIVIKITDQGYVTSRSHGHHYNGKVPYDA 98
      |||||

RESULT 8
D98004
histidine motif-containing protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zock, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <SUR>
A:Cross-references: GB:AE007317; PIDN:AAK93864.1; PID:gl5458682; GSPDB:GN00174
C:Genetics:
A:Gene: bnpA
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match          3.1%; Score 32; DB 2; Length 855;
Best Local Similarity 100.0%; Pred. No. 2.6e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 AEQIVIKITDQGYVTSRSHGHHYNGKVPYDA 98
      |||||
Db 77 AEQIVIKITDQGYVTSRSHGHHYNGKVPYDA 108
      |||||

RESULT 9
T46758
hypothetical 92.4K protein - Streptococcus agalactiae
C:Species: Streptococcus agalactiae
C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C:Accession: T46758
R:Spellerberg, B.; Rozdzinski, E.; Martin, S.; Weber-Heymann, J.; Schnitzler, N.; Luet
Infect. Immun. 67, 871-878, 1999
A:Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attachmen
A:Reference number: 224091; MUID:99115568; PMID:9916102
A:Accession: T46758
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-822 <SPE>
A:Cross-references: EMBL:AF062533; NID:g4249622; PIDN:RAD13797.1; PID:g4249624
A:Experimental source: strain R268
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match          1.5%; Score 16; DB 2; Length 822;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 KITDQGYVTSRSHGHHYH 88
      |||||
Db 75 KITDQGYVTSRSHGHHYH 90
      |||||

RESULT 10
E97985
hypothetical protein spr0909 [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: E97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zock, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-46 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99713.1; PID:gl5458516; GSPDB:GN00174
C:Genetics:
A:Gene: spr0909

```


C:Date: 1-92 <KLE>
A:Cross-references: GB:AE000943; GB:AE000782; MID:g2689266; PIDN:AA88923.1; PID:g264818

Query Match 0.8%; Score 8; DB 2; Length 92;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 843 VEKELSE 850
Db 78 VEKELSE 85
|||||

RESULT 13
A61291
ferredoxin [2Fe-2S] - parsley
C:Species: Petroselinum crispum (parsley)
C:Date: 27-Jun-1994 #sequence_revision 19-May-1995 #text_change 07-May-1999
C:Accession: A61291
R:Nakano, T.; Hase, T.; Matsubara, H.
J. Biochem. 90, 1725-1730, 1981
A:Title: The complete amino acid sequence of parsley (Petroselinum sativum) ferredoxin.
A:Reference number: A61291; MUID:82142254; PMID:7334006
A:Accession: A61291
A:Molecule type: protein
A:Residues: 1-96 <NAK>
A:Note: the source is designated as Petroselinum sativum
C:Superfamily: ferredoxin [2Fe-2S]; ferredoxin [2Fe-2S] homology
C:Keywords: 2Fe-2S; electron transfer; iron-sulfur protein; metalloprotein
F:24-78/Domain: ferredoxin [2Fe-2S] homology <FER>
F:39,44,47,77/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 0.8%; Score 8; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 5 VKLITPDG 12
|||||

RESULT 14
FEED
ferredoxin [2Fe-2S] - European elder (tentative sequence)
C:Species: Sambucus nigra (European elder)
C:Date: 30-Apr-1980 #sequence_revision 30-Apr-1980 #text_change 31-Mar-2000
C:Accession: A00233
R:Takurai, I.A.H.; Boulter, D.
Phytochemistry 18, 1481-1484, 1979
A:Title: The amino acid sequence of ferredoxin from Sambucus nigra.
A:Reference number: A00233
A:Accession: A00233
A:Molecule type: protein
A:Residues: 1-97 <TAK>
A:Note: the amidation states of residues 57, 58, 60, 61, 68, and 88 were identified by H
C:Superfamily: ferredoxin [2Fe-2S]; ferredoxin [2Fe-2S] homology
C:Keywords: 2Fe-2S; electron transfer; iron-sulfur protein; metalloprotein
F:1-97/Product: ferredoxin [2Fe-2S] #status experimental <MAT>
F:24-78/Domain: ferredoxin [2Fe-2S] homology <FER>
F:39,44,47,77/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 0.8%; Score 8; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 5 VKLITPDG 12
|||||

RESULT 15
FEYB6
ferredoxin [2Fe-2S] [validated] - Synechocystis sp.
C:Species: Synechocystis sp.

C:Date: 13-Jun-1983 #sequence_revision 13-Nov-1998 #text_change 15-Sep-2000
A:Accession: S76345; A56811; A00247
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s

A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76345
A:Molecule type: DNA
A:Residues: 1-97 <KAN>
A:Cross-references: EMBL:D64000; GB:AB001339; NID:g1001484; PIDN:BA010157.1; PID:g100
A:Experimental source: strain PCC 6803
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
R:Bottin, H.; Lagoutte, B.
Biochim. Biophys. Acta 1101, 48-56, 1992
A:Title: Ferredoxin and flavodoxin from the cyanobacterium Synechocystis sp PCC 6803.
A:Reference number: A56811; MUID:92338182; PMID:1633177
A:Accession: A56811
A:Molecule type: protein
A:Residues: 2-97 <BOT>
A:Experimental source: strain PCC 6714
A:Note: sequence extracted from NCBI backbone (NCBI:P109680)
R:Hase, T.; Inoue, K.; Matsubara, H.; Williams, M.M.; Rogers, L.J.
J. Biochem. 92, 1357-1362, 1982
A:Title: Amino acid sequence of Synechocystis 6714 ferredoxin: a unique structural fe
A:Reference number: A00247; MUID:83108768; PMID:6818221
A:Accession: A00247
A:Molecule type: protein
A:Residues: 2-14, N, 16-97 <HAS>
A:Experimental source: strain PCC 6714
R:Lelong, C.; Setif, P.; Bottin, H.; Andre, F.; Neumann, J.M.
submitted to the Brookhaven Protein Data Bank, September 1995
A:Reference number: A65450; PDB:IDOX
A:Contents: annotation; conformation and disulfide bond assignments by (1)H- and (15)
R:Lelong, C.; Setif, P.; Bottin, H.; Andre, F.; Neumann, J.M.
submitted to the Brookhaven Protein Data Bank, September 1995
A:Reference number: A65451; PDB:IDOX
A:Contents: annotation; conformation and disulfide bond assignments by (1)H- and (15)
R:Lelong, C.; Setif, P.; Bottin, H.; Andre, F.; Neumann, J.M.
Biochemistry 34, 14462-14473, 1995
A:Title: (1)H and (15)N NMR sequential assignment, secondary structure, and tertiary
A:Reference number: A58608; MUID:96062510; PMID:7578051
C:Superfamily: ferredoxin [2Fe-2S]; ferredoxin [2Fe-2S] homology
C:Keywords: 2Fe-2S; electron transfer; iron-sulfur protein; metalloprotein
F:2-97/Product: ferredoxin [2Fe-2S] #status experimental <MAT>
F:25-79/Domain: ferredoxin [2Fe-2S] homology <FER>
F:19-86/Disulfide bonds: #status predicted
F:40,45,48,78/Binding site: 2Fe-2S cluster (Cys) (covalent) #status experimental

Query Match 0.8%; Score 8; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 6 VKLITPDG 13
|||||

RESULT 16
S39689
vwcd protein - Bacillus subtilis
N:Alternate names: hypothetical protein ipa-34d
C:Species: Bacillus subtilis
C:Date: 07-Oct-1994 #sequence_rev#sion 26-May-1995 #text_change 15-Oct-1999
C:Accession: S39689; G70052
R:Glaser, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionesc
A.; Rapoport, G.; Danchin, A.
Mol. Microbiol. 10, 371-384, 1993
A:Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region
A:Reference number: S39655; MUID:95020537; PMID:7934828
A:Accession: S39689

```

A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-127 <GLA>
A:Cross-references: EMBL:X73124; NID:g413923; PIDN:CAA51590.1; PID:g413958
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertozzi, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chodura, A.; Enrich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.F.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A95580; MUID:98044033; PMID:9384377
A:Accession: G70052
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-127 <KUN>
A:Cross-references: GB:Z99123; GB:AL009126; NID:g2636240; PIDN:CAB15847.1; PID:ell86320;
A:Experimental source: strain 168
C:Genetics:
A:Gene: ywcd
C:Keywords: transmembrane protein

Query Match 0.8%; Score 8; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 LOSLUKEL 275
DB 59 LOSLUKEL 66

RESULT 17
PERZ
ferradoxin [2Fe-2S] I precursor - rice
C:Species: Oryza sativa (rice)
C>Date: 31-Mar-1989 #sequence_revision 28-May-1999 #text_change 16-Jun-2000
C:Accession: T03738; S03730; JTO223
R:Ohmori, K.; Doyama, N.; Ida, S.
Plant Physiol. 111, 348, 1996
A>Title: Molecular cloning of a rice leaf ferradoxin cDNA.
A:Reference number: Z15043
A:Accession: T03738
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-139 <OHM>
A:Cross-references: EMBL:D30763; PIDN:BA06436.1
A:Experimental source: tsugita japonica, cv. Kinmaze
R:Kamo, M.; Kotani, N.; Tsugita, A.; He, Y.K.; Nozu, Y.
Protein Seq. Data Anal. 2, 289-293, 1989
A>Title: Amino acid sequences of ferradoxins from rice cultivars, japonica and indica.
A:Reference number: S03730; MUID:89367259; PMID:2771933
A:Accession: S03730
A:Molecule type: protein
A:Residues: 44-139 <KAM>
A>Note: sequences from cultivars japonica and indica are identical
C:Superfamily: ferradoxin [2Fe-2S]; ferradoxin [2Fe-2S] homology
C:Keywords: 2Fe-2S; electron transfer; iron-sulfur protein; metalloprotein
F:44-139/Product: ferradoxin [2Fe-2S] I #status experimental <MAT>
F:67-121/Domain: ferradoxin [2Fe-2S] homology <PER>
F:82,87,90,120/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 0.8%; Score 8; DB 1; Length 139;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-127 <GLA>
A:Cross-references: EMBL:X73124; NID:g413923; PIDN:CAA51590.1; PID:g413958
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertozzi, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chodura, A.; Enrich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.F.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A95580; MUID:98044033; PMID:9384377
A:Accession: G70052
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-127 <KUN>
A:Cross-references: GB:Z99123; GB:AL009126; NID:g2636240; PIDN:CAB15847.1; PID:ell86320;
A:Experimental source: strain 168
C:Genetics:
A:Gene: ywcd
C:Keywords: transmembrane protein

Query Match 0.8%; Score 8; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
DB 58 NGVAIPHG 55

RESULT 19
AD3475
protein-Npi-phosphohistidine-sugar phosphotransferase (EC 2.7.1.65) [imported] - Bruc
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: AD3475
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov, M.; Mazur, M.; Goldstein, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let, proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3475
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <KUR>
A:Cross-references: GB:AE008917; PIDN:AA152967.1; PID:g17983818; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11786
A:Map position: 1
C:Superfamily: phosphotransferase system enzyme II; phosphotransferase system mannito
C:Keywords: phosphotransferase

Query Match 0.8%; Score 8; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
DB 60 NGVAIPHG 67

RESULT 20
A69349
conserved hypothetical transmembrane protein AF0793 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

```

C:Accession: A69349
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiaich, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: A69349
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-207 <KLE>
A:Cross-references: GB:AE001050; GB:AE000782; NID:g2689373; PIDN:AAB90448.1; PID:g264981

Query Match 0.8%; Score 8; DB 2; Length 207;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 LVEEQIKA 480
|||||||
Db 58 LVEEQIKA 65

RESULT 21
RGRGLT
ribosomal protein L1 - Thermotoga maritima (strain MSB8)
C:Species: thermotoga maritima
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 21-Jul-2000
C:Accession: C44466; H72375; S19900
R:Liao, D.; Dennis, P.P.
J. Biol. Chem. 267, 22787-22797, 1992
A:Title: The organization and expression of essential transcription translation componen
A:Reference number: A44466; MUID:93054590; PMID:1429627
A:Accession: C44466
A:Molecule type: DNA
A:Residues: 1-233 <LIA>
A:Cross-references: EMBL:Z11839; NID:g407020; PIDN:CAA77860.1; PID:g48185
A:Note: sequence extracted from NCBI backbone (NCBIP:118056)
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: H72375
A:Molecule type: DNA
A:Residues: 1-233 <ARN>
A:Cross-references: GB:AE001723; GB:AE000512; NID:g4980953; PIDN:AAD35538.1; PID:g498096
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0455
C:Superfamily: Escherichia coli ribosomal protein L1
C:Keywords: protein biosynthesis; ribosome; RNA binding

Query Match 0.8%; Score 8; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 267 NLQSLIKE 274
|||||||
Db 226 NLQSLIKE 233

RESULT 22
RGR22
ribosomal protein L2 - rice chloroplast
C:Species: chloroplast Oryza sativa (rice)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 17-Feb-1995
C:Accession: JQ0270; S05150; JA0093
R:Shimada, H.; Whittier, R.F.; Hiratsuka, J.; Macda, Y.; Hirai, A.; Sugiyura, M.
submitted to JIPID, December 1989
A:Reference number: JQ0200

A:Accession: JQ0270
A:Molecule type: DNA
A:Residues: 1-273 <SHI>
A:Experimental source: cv. Nihonbare
R:Hiratsuka, J.; Shimada, H.; Whittier, R.; Ishibashi, T.; Sakamoto, M.; Mori, M.; Ko
M.
Mol. Gen. Genet. 217, 185-194, 1989
A:Title: The complete sequence of the rice (Oryza sativa) chloroplast genome: intermo
of the cereals.
A:Reference number: S05080; MUID:89364698; PMID:2770692
A:Accession: S05150
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-273 <HIR>
A:Experimental source: cv. Nihonbare
A:Note: this sequence was submitted to EMBL, July 1989
R:Moore, E.; Wu, R.
Gene 70, 1-12, 1988
A:Title: Organization and nucleotide sequence of genes at both junctions between the
A:Reference number: JA0092; MUID:89196901; PMID:3240862
A:Accession: JA0093
A:Molecule type: DNA
A:Residues: 1-68, 'E', 70-199, 'K', 201-229, 'MGAVKGKPLVEXNPQ', 247-273 <MOO>
C:Genetics:
A:Gene: rpl2
A:Genome: chloroplast
A:Start codon: ACG
A:Introns: 131/1
C:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 0.8%; Score 8; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1022 LRLPSGEV 1029
|||||||
Db 172 LRLPSGEV 179

RESULT 23
R52M2
ribosomal protein L2 - maize chloroplast
C:Species: Chloroplast Zea mays (maize)
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 21-Jul-2000
C:Accession: S10500; S17874; S58639; S58595
R:Kavousi, M.; Giese, K.; Larrinua, I.M.; McLaughlin, W.E.; Subramanian, A.R.
Nucleic Acids Res. 18, 4244, 1990
A:Title: Nucleotide sequence and map positions of the duplicated gene for maize (Zea
A:Reference number: S10500; MUID:90332419; PMID:2377464
A:Accession: S10500
A:Molecule type: DNA
A:Residues: 1-273 <KAV>
A:Cross-references: EMBL:X53066
R:Hoch, B.; Maier, R.M.; Appel, K.; Igloi, G.L.; Koessel, H.
Nature 353, 178-180, 1991
A:Title: Editing of a chloroplast mRNA by creation of an initiation codon.
A:Reference number: S17874; MUID:91367263; PMID:1653905
A:Accession: S17874
A:Molecule type: mRNA
A:Residues: 1-150 <HOC>
A:Cross-references: EMBL:X62070
R:Maier, R.M.; Neckermann, K.; Igloi, G.L.; Koessel, H.
J. Mol. Biol. 251, 614-628, 1995
A:Title: Complete sequence of the maize chloroplast genome: gene content, hotspots of
A:Reference number: S58531; MUID:95395841; PMID:7666415
A:Accession: S58639
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-273 <MAI>
A:Cross-references: EMBL:X86563; NID:g902200; PIDN:CAA60329.1; PID:g2673858
A:Genetics: GENI
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995

A:Accession: S58595
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-273 <WAK>
A:Cross-references: EMBL:X86563; NID:9022200; PIDN:CAA50329.1; PID:g2673858
A:Genetics: GEN2
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics: <GEN1>
A:Map position: IR(A)
A:Genome: chloroplast
A:Start codon: ACG
A:Introns: 130/3
A:Note: the start codon AUG results from mRNA editing of ACG
C:Genetics: <GEN2>
A:Map position: IR(B)
A:Genome: chloroplast
A:Introns: 130/3
C:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; protein biosynthesis; ribosome; RNA editing

Query Match 0.8%; Score 8; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
Db 172 LRLPSGEV 179
|||||

RESULT 24
T11810
ribosomal protein L2 - Norway spruce chloroplast
C:Species: chloroplast Picea abies (Norway spruce)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T11810
R:Kluemper, S.; Kanka, S.; Riesner, D.; Etscheid, M.
submitted to the EMBL Data Library, March 1997
A:Description: Characterisation of a Norway spruce chloroplast DNA clone: Complete nucleotide sequence
A:Reference number: Z17349
A:Accession: T11810
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-275 <KLU>
A:Cross-references: EMBL:U92462; NID:g2959581; PID:g2959586
C:Genetics:
A:Gene: rpl2
A:Genome: chloroplast
A:Introns: 133/1
C:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; protein biosynthesis; ribosome

Query Match 0.8%; Score 8; DB 2; Length 275;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
Db 174 LRLPSGEV 181
|||||

RESULT 25
T07361
ribosomal protein L2 - Chlorella vulgaris chloroplast
C:Species: chloroplast Chlorella vulgaris
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C:Accession: T07361
R:Wakasugi, T.; Nagai, T.; Kapoor, M.; Sugita, M.; Ito, S.; Tsudzuki, J.; Nakas
Proc. Natl. Acad. Sci. U.S.A. 94, 5967-5972, 1997
A:Title: Complete nucleotide sequence of the chloroplast genome from the green alga Chlorella vulgaris
A:Reference number: Z15985; MUID:97303241; PMID:9159184
A:Accession: T07361
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-275 <WAK>
A:Cross-references: EMBL:AB001684; NID:g2224352; PIDN:BAA58009.1; PID:g2224525
C:Genetics:
A:Gene: rpl2
A:Genome: chloroplast
A:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; ribosome

Query Match 0.8%; Score 8; DB 2; Length 275;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
Db 174 LRLPSGEV 181
|||||

RESULT 26
T07531
ribosomal protein L2 - Japanese black pine chloroplast (fragment)
C:Species: chloroplast Pinus thunbergiana (Japanese black pine)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000
C:Accession: T07531
R:Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugitara, M.
Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
A:Title: Loss of all ndh genes as determined by sequencing the entire chloroplast gen
A:Reference number: Z16030; MUID:95024047; PMID:7937893
A:Accession: T07531
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-276 <WAK>
A:Cross-references: EMBL:DI7510; NID:g529643; PIDN:BAA23474.1; PID:g2626947
C:Genetics:
A:Gene: rpl2
A:Genome: chloroplast
A:Note: Intron positions not resolved (incomplete sequence)
C:Superfamily: Escherichia coli ribosomal protein L2
C:Keywords: chloroplast; ribosome

Query Match 0.8%; Score 8; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
Db 175 LRLPSGEV 182
|||||

RESULT 27
AF2987
ABC transporter, membrane spanning protein Atu3503 [imported] - Agrobacterium tumefac
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AF2987
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Ram
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF2987
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-277 <KUR>
A:Cross-references: GB:AB008689; PIDN:AAL44316.1; PID:g17741906; GSPDB:GN00187
C:Genetics:
A:Gene: Atu3503
A:Map position: linear chromosome
C:Superfamily: maltose transport protein malG

```

Query Match          0.8%; Score 8; DB 2; Length 277;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 981 LDPALPEA 988
    |||||
Db 166 LDPALPEA 173

RESULT 28
B98296
sulfate ABC transporter, permease protein VC0539 [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: B98296
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blarchard, M.; Quorillo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome sequence of the plant pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: B98296
A:Status: preliminary
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-278 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89892.1; PID:g15159839; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L2652
A:Map position: linear chromosome
C:Superfamily: maltose transport protein malG

Query Match          0.8%; Score 8; DB 2; Length:278;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 981 LDPALPEA 988
    |||||
Db 167 LDPALPEA 174

RESULT 29
AI3633
hypothetical protein BMEII0994 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AI3633
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AI3633
A:Status: preliminary
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-278 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL54236.1; PID:g17985208; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEII0994
A:Map position: II

Query Match          0.8%; Score 8; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 302 PNGVAIPH 309
    |||||
Db 258 PNGVAIPH 265

RESULT 30
H59094
hypothetical protein pX01-32 - Bacillus anthracis virulence plasmid pX01
C:Species: Bacillus anthracis

```

```

C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C:Accession: H59094
R:Oknaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koeh
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harb
A:Reference number: A59091; MUID:99445483; PMID:10515943
A:Accession: H59094
A:Status: preliminary
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <OXI>
A:Cross-references: GB:AF065404; NID:94894216; PIDN:AAD32336.1; PID:g4894248
C:Genetics:
C:Experimental source: strain Sterne
C:Genetics:
A:Gene: pX01-32
A:Genome: plasmid
C:Superfamily: Bacillus anthracis virulence plasmid pX01 hypothetical protein pX01-32

Query Match          0.8%; Score 8; DB 2; Length 327;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 587 EGNKVYTG 594
    |||||
Db 154 EGNKVYTG 161

RESULT 31
G83148
DNA polymerase III, delta subunit PA3989 [imported] - Pseudomonas aeruginosa (strain
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83148
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: G83148
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-345 <STO>
A:Cross-references: GB:AE004816; GB:AE004091; NID:g9950168; PIDN:AAG07376.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: hola; PA3989

Query Match          0.8%; Score 8; DB 2; Length 345;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1020 IELRLPSG 1027
    |||||
Db 79 IELRLPSG 86

RESULT 32
AD2087
gas vesicle protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2087
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuriitz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2087
A:Status: preliminary
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-410 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAE73950.1; PID:g17131342; GSPDB:GN00179

```

A:Experimental source: strain PCC 7120

C:Genetics:
A:Gene: gvpN

Query Match 0.8% Score 8; DB 2; Length 410;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 321 LSALEEKI 328
DB 145 LSALEEKI 152

RESULT 33

F81329
serine proteinase (proteinase DO) (EC 3.4.21.-) Cj1228c [imported] - Campylobacter jejuni

C:Species: Campylobacter jejuni

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: F81329
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, W.P.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrett, N.; et al. 2000. Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervirulence

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: F81329

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-472 <PAR>

A:Cross-references: GB:AL139077; GB:AL111168; NID:g6968444; PIDN:CAB73482.1; PID:g696868

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: htrA; Cj1228c

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase, serine proteinase

Query Match 0.8% Score 8; DB 2; Length 472;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 EVVSSSLGS 356
DB 98 EVVSSSLGS 105

RESULT 34

B83735

cassette chromosome recombinase B1 BH0682 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: B83735

R:Ikemura, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai, T.; et al. 2000. Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and its implications for the evolution of the Bacteroidetes phylum

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: B83735

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-522 <STO>

A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA004401.1; GSPDB:GN000004

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0682

Query Match 0.8% Score 8; DB 2; Length 522;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 829 KLDKVEE 836
DB 203 KLDKVEE 210

RESULT 35

B97331

membrane associated methyl-accepting chemotaxis protein with HAMP domain [imported]

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: B97331

R:Nolling, J.; Boret, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L. J. Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97331

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-573 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK81437.1; PID:g15026604; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3511

Query Match 0.8% Score 8; DB 2; Length 573;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 NVVNLKKN 605
DB 259 NVVNLKKN 266

RESULT 36

S61157

hypothetical protein YDR362c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein D9476.13

C:Species: Saccharomyces cerevisiae

C>Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 19-Apr-2002

C:Accession: S61157

R:Du, Z.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of S. cerevisiae cosmid 9476.

A:Reference number: S61148

A:Accession: S61157

A:Molecule type: DNA

A:Residues: 1-672 <DU2>

A:Cross-references: EMBL:U28372; NID:g849170; PID:g849183; GSPDB:GN05004; MIPS:YDR362

C:Genetics:

A:Gene: SGD:YDR362C

A:Cross-references: SGD:S0002770

A:Map position: 4R

C:Superfamily: Saccharomyces cerevisiae hypothetical protein YDR362C

Query Match 0.8% Score 8; DB 2; Length 672;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 YIFNPKDI 384
DB 492 YIFNPKDI 499

RESULT 37

S48261

hypothetical protein YBR094w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YBR0821

C:Species: Saccharomyces cerevisiae

C>Date: 01-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 13-Apr-2002

C:Accession: S48261; S45962; S41800; S44676

R:Mannhaupt, G.; Stucka, R.; Ehle, S.; Vetter, I.; Feldmann, H.

Yeast 10, 1363-1381, 1994

A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.

A:Reference number: S48255; MUID:95208357; PMID:7900426

A:Accession: S48261

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-753 <MAN>

A:Cross-references: EMBL:X78993; NID:g476045; PIDN:CAA55599.1; PID:g476052
R:Feldmann, H.; Manhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45927
A:Accession: S45962
A:Molecule type: DNA
A:Residues: 1-753 <FE2>
A:Cross-references: EMBL:Z35963; NID:g536366; PIDN:CAA85047.1; PID:g536367; MIPS:YBR094W
R:Decker, P.J.T.; Hoekert, W.; van Oosterum, K.; Grivell, L.A.
submitted to the EMBL Data Library, December 1992
A:Reference number: S31313
A:Accession: S41800
A:Molecule type: DNA
A:Residues: 167-351, 'TPE', 356-449, 'A', 451-562, 'R', 564-753 <DEK>
A:Cross-references: EMBL:X69881; NID:g4086; PIDN:CAA49508.1; PID:g4088
C:Genetics:
A:Cross-references: SGD:S0000298
A:Map position: 2R

Query Match 0.8%; Score 8; DB 2; Length 753;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 KAHLACKN 233
Db 50 KAHLACKN 57
|||||||

RESULT 38
T18280
hypothetical protein D2 - slime mold (Dictyostellium discoideum)
C:Species: Dictyostellium discoideum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
A:Accession: T18280
R:Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh
Genetics 148, 1117-1125, 1998
A:Title: Dictyostellium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1
A:Reference number: Z14684; MUID:98198836; PMID:9539429
A:Accession: T18280
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-763 <RIE>
A:Cross-references: EMBL:U00796; NID:g2702254; PID:g2702255; PIDN:AAC18631.1

Query Match 0.8%; Score 8; DB 2; Length 763;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 NMLVKLIT 645
Db 632 NMLVKLIT 639
|||||||

RESULT 39
T40674
protein transport protein sec23 homolog - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
A:Accession: T40674
R:Lynne, M.; Rajandream, M.A.; Barrell, B.G.; Wedler, H.; Kutzner, M.; Wambutt, R.
submitted to the EMBL Data Library, January 1999
A:Reference number: Z21944
A:Accession: T40674
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-765 <LYN>
A:Cross-references: EMBL:AL035263; PIDN:CAA28877.1; GSPDB:GN00067; SPDB:SPBC776.04
A:Experimental source: strain 972h-; cosmid c776
C:Genetics:
A:Gene: SPDB:SPBC776.04
A:Map position: 2
A:Introns: 11/1; 68/2; 79/1
C:Superfamily: SEC23 protein

Query Match 0.8%; Score 8; DB 2; Length 765;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 AVIVLSLSL 20
Db 144 AVIVLSLSL 151
|||||||

RESULT 40
G96494
protein F7F22.4 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
A:Accession: G96494
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G96494
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-813 <STO>
A:Cross-references: GB:AE005173; NID:g6691183; PIDN:AAF24521.1; GSPDB:GN00141
C:Genetics:
A:Gene: F7F22.4
A:Map position: 1

Query Match 0.8%; Score 8; DB 2; Length 813;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEQIK 479
Db 432 DLTEQIK 439
|||||||

RESULT 41
C75313
probable proteinase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
A:Accession: C75313
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: C75313
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-840 <WHI>
A:Cross-references: GB:AF002047; GB:AE000513; NID:g6459915; PIDN:AAF11680.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2130
A:Map position: 1

Query Match 0.8%; Score 8; DB 2; Length 840;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 DGLVFDPA 295
|||||||

```

Db 332 DGLVDPDA 339
RESULT 42
G86393
protein T24P13.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86393
R:Theologidis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso-
ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cready, I.H.; Dewar, K.;
Nature 408, 816-820, 2000
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-984 <STO>
A:Cross-references: GB:AE005172; NID:g295721; PIDN:AAF87027.1; GSPDB:GN00141
C:Genetics:
A:Gene: T24P13.6
A:Map position: 1

Query Match 0.8%; Score 8; DB 2; Length 984;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 SLSSNPSS 363
Db 843 SLSSNPSS 850
|||||||

RESULT 43
S72635
exo-poly-alpha-galacturonosidase (EC 3.2.1.82) precursor - Thermoanaerobacterium thermos
C:Species: Thermoanaerobacterium thermosulfurigenes
C:Date: 29-Jul-1997 #sequence_revision 29-Jul-1997 #text_change 15-Oct-1999
C:Accession: S72635; S72621
R:Matuschek, M.; Sam, K.; Bahl, H.
submitted to the EMBL Data Library, March 1996
A:Description: Characterization of genes from Thermoanaerobacterium thermosulfurigenes
A:Reference number: S72635
A:Accession: S72635
A:Molecule type: DNA
A:Residues: 1-1148 <MATU>
A:Cross-references: EMBL:U50951; NID:g1542972; PIDN:AAB08040.1; PID:g1255234
R:Matuschek, M.; Sam, K.; Zibat, A.; Bahl, H.
Mol. Gen. Genet. 252, 493-496, 1996
A:Title: Characterization of genes from Thermoanaerobacterium thermosulfurigenes E1 tha
A:Reference number: S72621; MUID:97033555; PMID:8879252
A:Accession: S72621
A:Molecule type: DNA
A:Residues: 729-1148 <MAW>
A:Cross-references: EMBL:U50951
A:Experimental source: EM1
C:Genetics:
A:Gene: palA
C:Superfamily: Thermoanaerobacterium endo-1,4-beta-xylanase A; S-layer repeat homology;
a xylanase A cellulose-binding repeat homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1148/Product: exo-poly-alpha-galacturonosidase #status predicted <MAT>
F:959-1022/Domain: S-layer repeat homology <SLR1>
F:1028-1081/Domain: S-layer repeat homology <SLR2>
F:1092-1144/Domain: S-layer repeat homology <SLR3>

Query Match 0.8%; Score 8; DB 2; Length 1148;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 LKNSTFNN 611
Db 598 LKNSTFNN 605
|||||||

RESULT 44
T14265
golgin-245 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14265
R:Gay, D.; Beller, B.M.; Zhao, H.; Bhandoola, A.; Tomayko, M.M.; Murali, R.; Marks, M.
submitted to the EMBL Data Library, February 1998
A:Description: Olp-1, the murine golgin-245/p230 homologue, is upregulated during oli
A:Reference number: Z17947
A:Accession: T14265
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1218 <GAY>
A:Cross-references: EMBL:AF051357; NID:g2952521; PID:g2952522; PIDN:AAC05573.1
A:Experimental source: strain BALB/c; brain
C:Genetics:
A:Gene: olp-1

Query Match 0.8%; Score 8; DB 2; Length 1218;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 SAVIVSLS 19
Db 114 SAVIVSLS 121
|||||||

RESULT 45
T45070
protein kinase homolog R31240.1 [imported] - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45070
R:Lamerdin, J.; McCready, P.; Stilwagen, S.; Ramirez, M.; Carrano, A.
submitted to the EMBL Data Library, November 1996
A:Description: Characterization by genomic sequence analysis of a gene-rich 111 kb re
A:Reference number: Z22906
A:Accession: T45070
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1237 <LAM>
A:Cross-references: EMBL:AD000092; PIDN:AAB51171.1
A:Experimental source: cell line 5HL2-B; fibroblast
C:Genetics:
A:Map position: 19p13.2
A:Note: Intron positions not resolved (incomplete sequence)

Query Match 0.8%; Score 8; DB 2; Length 1237;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 SLSSNPSS 363
Db 595 SLSSNPSS 602
|||||||

Search completed: May 13, 2003, 14:01:41
Job time : 35 secs

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: May 13, 2003, 13:56:37 ; Search time 16 Seconds
(without alignments)
2693.369 Million cell updates/sec
Title: US-09-471-255-2
Perfect score: 1039
Sequence: 1 MKFSKKYIAGSAVIVSL.....IELPLPGEVKKMLSDFIA 1039
Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 112892 seqs, 41476328 residues
Word size : 8
Total number of hits satisfying chosen parameters: 15
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 500 summaries
Database : SwissProt_40.*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	0.8	92	1 YN31_ARCFU	Q27953 archaeoglob
2	8	0.8	96	1 FER1_ORYSA	P11051 oryza sativ
3	8	0.8	96	1 FER_SIN4	P27320 synechocyst
4	8	0.8	96	1 FER_SIN4	P00243 synechocyst
5	8	0.8	97	1 FER_SAMN1	P00226 sambucus ni
6	8	0.8	127	1 YWCD_BAGSU	P39602 bacillus su
7	8	0.8	233	1 RL1_THENA	P29393 thermotoga
8	8	0.8	273	1 RK2_HORVU	P41086 hordeum vul
9	8	0.8	273	1 RK2_MAIZE	P17788 zea mays (m
10	8	0.8	273	1 RK2_ORYSA	P17351 oryza sativ
11	8	0.8	275	1 RK2_CHLYU	P56367 chlorella v
12	8	0.8	275	1 RK2_PICAB	O62954 picea abies
13	8	0.8	276	1 RK2_PINTH	O62940 pinus thunb
14	8	0.8	335	1 GVPN_ANAFL	P55150 anabaena fl
15	8	0.8	753	1 YBU4_YEAST	P38234 saccharomyc

ALIGNMENTS

RESULT 1
ID YN31_ARCFU STANDARD; PRT: 92 AA.
AC Q27953;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF2331.
GN AF2331.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;

[1]
SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; Pubmed=9389475;
RA Kleck H. P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
Rechum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D.,
Richardson D.L., Kerlavage A.R., Graham D.E., Kyrides N.C.,
Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback L.,
Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997)
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000943; AAB88923.1;
DR TIGR; AF2331;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 92 AA; 10521 MW; CCC444D2796013CC CRC64;
Query Match 0.88; Score 8; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 843 VEKEKLSK 850
DB 78 VEKEKLSK 85
RESULT 2
ID FER1_ORYSA STANDARD; PRT: 96 AA.
AC P11051;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferredoxin I.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE.
RC STRAIN=cv. Japonica, and cv. Indica;
RX MEDLINE=89367259; Pubmed=2771933;
RA Kamo M., Kotani N., Tsugita A., He Y.K., Nozu Y.;
RT "Amino acid sequences of ferredoxins from rice cultivars, Japonica
RT and Indica";
RL Protein Seq. Data Anal. 2:289-293(1989).
CC -!- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
DR PIR; J02023; FERZ
DR PIR; S03730; S03730.
DR HSSP; P00221; 1A70.
DR InterPro; IPR000564; 2Fe2S_ferredoxin.
DR InterPro; IPR001041; Ferredoxin.
DR Pfam; PF00111; fer2; 1.
DR PRINTS; PR00159; 2Fe2SFRDXIN.

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: D85607; BAA24020.1; -
DR EMBL: U38802; AAB72025.1; -
DR EMBL: D64000; BAA10197.1; -
DR PDB: 1DOX; 08-MAR-96.
DR PDB: 1DOY; 08-MAR-96.
DR InterPro: IPR000584; 2FE2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SFRDXIN.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Electron transport; Iron-sulfur; 3D-structure; Complete proteome.
FT INIT_MET 0
FT METAL 39 39 IRON-SULFUR (2FE-2S).
FT METAL 44 44 IRON-SULFUR (2FE-2S).
FT METAL 47 47 IRON-SULFUR (2FE-2S).
FT METAL 77 77 IRON-SULFUR (2FE-2S).
FT DISULFID 18 85 PROBABLE.
SQ SEQUENCE 96 AA; 10232 MW; 73866DD68E369451 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
DB 5 VKLITPDG 12

RESULT 4
FER_SYNY4
ID_FER_SYNY4 STANDARD; PRT; 96 AA.
AC P00243; 1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1988 (Rel. 09, Last annotation update)
DE Ferredoxin.
OS Synechocystis sp. (strain PCC 6714).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1147;
RN [1]
RP SEQUENCE.
RX MEDLINE=83108768; PubMed=6818221;
RA Hase T., Inoue K., Matsubara H., Williams M.M., Rogers L.J.;
RT "Amino acid sequence of Synechocystis 6714 ferredoxin: a unique
RT structural feature of unicellular blue-green algal ferredoxin."
RL J. Biochem. 92:1357-1362(1982).
CC -!- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER.
CC PIR: A00247; FEYB6.
DR HSP: P27320; 1DOX.
DR InterPro: IPR000584; 2FE2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SFRDXIN.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Electron transport; Iron-sulfur.
FT METAL 39 39 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 44 44 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 47 47 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 77 77 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 96 AA; 10259 MW; 5E96781964135BBB CRC64;

Query Match 0.8%; Score 8; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648

RESULT 3
FER_SYNY3
ID_FER_SYNY3 STANDARD; PRT; 96 AA.
AC P27320;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferredoxin I.
GN PEIF OR FED OR SLL0011.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakamoto H., Suzuki T.;
RT "Cloning, characterization and transcriptional studies of ferredoxin
RT genes from the mesophilic cyanobacterium Synechocystis sp. PCC 6803
RT and the thermophilic cyanobacterium Synechococcus vulcanus."
RL Physiol. Plantarum 101:199-205(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE FROM N.A.
RA Cassier-Chauvat C., Poncelet M., Viljoing S., Chauvat F.;
RL Submitted (APR-1996) to the EMBL/GenBank/DDJB databases.
RN [4]
RP SEQUENCE.
RX MEDLINE=92338182; PubMed=1633177;
RA Bottin H., Lagoutte B.;
RT "Ferredoxin and flavodoxin from the cyanobacterium Synechocystis sp
RT PCC 6803."
RL Biochim. Biophys. Acta 1101:48-56(1992).
RN [5]
RP STRUCTURE BY NMR.
RX MEDLINE=96062510; PubMed=7578051;
RA Lelong C., Setif P., Bottin H., Andre F., Neumann J.-M.;
RT "1H and 15N NMR sequential assignment, secondary structure, and
RT tertiary fold of [2Fe-2S] ferredoxin from Synechocystis sp. PCC
RT 6803."
RL Biochemistry 34:14462-14473(1995).
CC -!- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -!- COFACTOR: BINDS 1 2FE-2S CLUSTER.
CC -!- MISCELLANEOUS: THE MIDPOINT REDOX POTENTIAL FOR THIS PROTEIN IS
CC -412 MV.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

Db      5 VKLITPDG 12
|||||
RESULT 5
FER_SAMMI
ID FER_SAMMI STANDARD; PRT; 97 AA.
AC P00226;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Ferredoxin.
OS Sambucus nigra (European elder).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Dipsacales; Adoxaceae; Sambucus.
OX NCBI_TaxID=4202;
RN [1]
RP SEQUENCE.
RA Takurai I.A.H., Boulter D.;
RA "The amino acid sequence of ferredoxin from Sambucus nigra.";
RA Phytochemistry 18:1481-1484(1979).
RT
CC -!- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -!- COPACITOR: BINDS 1 2FE-2S CLUSTER.
CC -!- SUBCELLULAR LOCATION: Chloroplast.
DR PIR: A00233; FEED.
DR HSSP: P00221; 1A70.
DR InterPro: IPR000564; 2Fe2s_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SPRDOXIN
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Electron transport; Iron-sulfur; Chloroplast.
FT METAL 39 39 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 44 44 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 47 47 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 77 77 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 97 AA; 10616 MW; D19C34A411C68AA2 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
|||||
Db      5 VKLITPDG 12
|||||

RESULT 6
YWCD_BACSU
ID YWCD_BACSU STANDARD; PRT; 127 AA.
AC P39602;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ywcd.
OS YWCD OR IPA-34D.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Tonescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Sanchana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caidwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goifeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwa A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Solido B.,
RA Sekiguchi J., Sekowska A., Serot S.J., Serrif P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein F., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE GTRA FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
DR EMBL: X73124; CAAS1590.1; -;
DR EMBL: Z99123; CAB15847.1; -;
DR PIR: S39689; S39689.
DR Subtilist; BG10580; ywcd.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 32 52 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
SQ SEQUENCE 127 AA; 14349 MW; 1A51311C0A0FA05F CRC64;

Query Match 0.8%; Score 8; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 LQSLKEL 275
|||||
Db      59 LQSLKEL 66
|||||

RESULT 7
RLA_THEMEA
ID RLI_THEMEA STANDARD; PRT; 233 AA.
AC P29393;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L1.

```



```

RN RNA EDITING OF INITIATOR CODON.
RX MEDLINE=91367263; PubMed=1653905;
RA Hoch B., Maier R.M., Appel K., Igloi G.L., Koessel H.;
RT "Editing of a chloroplast mRNA by creation of an initiation codon.";
RL Nature 353:176-180(1991)
CC -!- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC -!- CAUTION: THE INITIATOR METHIONINE IS CREATED BY RNA EDITING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53056; CA437241.1; ALT_SEQ.
DR EMBL; X86563; CAAG0329.1; ALT_SEQ.
DR EMBL; X86563; CAAG0371.1; ALT_SEQ.
DR EMBL; X62070; CA43983.1; ALT_SEQ.
DR PIR; S10500; R52N2.
DR PIR; S17874; S17874.
DR HSSP; P04257; IRL2.
DR MaizeDB; 66413; -.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW Ribosomal protein; Chloroplast; RNA editing.
RN RIBOSOMAL PROTEIN; Chloroplast; RNA editing.
SQ SEQUENCE 273 AA; 30065 MW; BA65197231EA3CA0 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1022 LRLPSGEV 1029
Db 172 LRLPSGEV 179
|||||||

RESULT 10
RK2_ORYSA STANDARD; PRT; 273 AA.
AC P17351; P92327;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chloroplast 50S ribosomal protein L2.
GN RPL2-A AND RPL2-B.
OS Oryza sativa (Rice).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89196901; PubMed=3240862;
RA Moon E., Wu R.;
RT "Organization and nucleotide sequence of genes at both junctions
RT between the two inverted repeats and the large single-copy region in
RT the rice chloroplast genome.";
RL Gene 70:1-12(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RX MEDLINE=89364698; PubMed=2770692;
RA Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M.,
RA Mori M., Kondo C., Honji Y., Sun C.-R., Meng B.-Y., Li Y.-Q.,
RA Kanno A., Nishizawa Y., Hirai A., Shinzaki K., Sugura M.;
RT "The complete sequence of the rice (Oryza sativa) chloroplast genome:
RT intermolecular recombination between distinct trna genes accounts for

```

```

RT a major plastid DNA inversion during the evolution of the cereals.";
RL Mol. Genet. 217:185-194(1989).
CC -!- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22826; -. NOT-ANNOTATED.CDS.
DR EMBL; L40578; AAD15254.1; ALT_SEQ.
DR EMBL; X15901; CA43924.1; ALT_SEQ.
DR EMBL; X15901; CA43928.1; ALT_SEQ.
DR PIR; JQ0270; R5R22.
DR HSSP; P04257; IRL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW Ribosomal protein; Chloroplast.
FT CONFLICT 69 69 Q -> E (IN REF. 1).
FT CONFLICT 200 200 Q -> K (IN REF. 1).
FT CONFLICT 230 246 HGGEGCKADIGRKKPTT -> GAVKGGKPLPEKNPQP
FT (IN REF. 1).
SQ SEQUENCE 273 AA; 29942 MW; AA153062AA8729B CRC64;

Query Match 0.8%; Score 8; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1022 LRLPSGEV 1029
Db 172 LRLPSGEV 179
|||||||

RESULT 11
RK2_CHLVU STANDARD; PRT; 275 AA.
AC P56367;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Chloroplast 50S ribosomal protein L2.
GN RPL2.
OS Chlorella vulgaris.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Chlorella.
OX NCBI_TaxID=3077;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TAM C-27 / Tamiya;
RX MEDLINE=97303241; PubMed=9159184;
RA Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
RA Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
RA Inamura A., Yoshinaga K., Sugura M.;
RT "Complete nucleotide sequence of the chloroplast genome from the
RT green alga Chlorella vulgaris: the existence of genes possibly
RT involved in chloroplast division.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
CC -!- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; AB001684; BAA58009.1;
DR HSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 275 AA; 30093 MW; 369FF92300441165 CRC64;

Query Match      0.8%; Score 8; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 174 LRLPSGEV 181

RESULT 12
RK2_PICAB STANDARD; PRT; 275 AA.
AC 062954;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Chloroplast 50S ribosomal protein L2.
GN RPL2.
OG Picea abies (Norway spruce) (Picea excelsa).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferales; Pinaceae; Picea.
OX NCBI_TaxID=3329;
RN [1]
RP SEQUENCE FROM N.A.
RA Kluepfer S., Kanka S., Riesner D., Etscheid M.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U92462; AAC95500.1;
DR HSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; FALSE_NEG.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 275 AA; 30180 MW; 1DFDE72CE798C378 CRC64;

Query Match      0.8%; Score 8; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 174 LRLPSGEV 181

RESULT 13
RK2_PINTH STANDARD; PRT; 276 AA.
AC 062940;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chloroplast 50S ribosomal protein L2.
GN RPL2.

```

```

OS Pinus thunbergii (Green pine) (Japanese black pine).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP SEQUENCE FROM N.A.
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugura M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii."
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -!- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D17510; BAA23474.1;
DR HSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRFAMs; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00467; RIBOSOMAL_L2; FALSE_NEG.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 276 AA; 30278 MW; 7F783A55998FB146 CRC64;

Query Match      0.8%; Score 8; DB 1; Length 276;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 175 LRLPSGEV 182

RESULT 14
GVPN_ANAFI STANDARD; PRT; 335 AA.
AC P55150;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gas vesicle protein gvpN.
GN GVPN.
OS Anabaena flos-aquae.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=1166;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCAP 1403/13f;
RX MEDLINE=97217585; PubMed=9063646;
RA Kinsman R., Hayes P.K.;
RT "Genes encoding proteins homologous to halobacterial Gyps N, J, K, F
RT & L are located downstream of gvpC in the cyanobacterium Anabaena
RT flos-aquae".
RL DNA Seq. 7:97-106(1997).
CC -!- SIMILARITY: BELONGS TO THE CBBO/NIRO/GVNO/GVNP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U17109; AAA58711.1;

```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:57:43 ; Search time 48 Seconds
(without alignments)
4460.062 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 1039

Sequence: 1 MKFSKKYTAGSAVIVLSL.....IELRPSGEVIKKNLSDFIA 1039

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 8

Total number of hits satisfying chosen parameters: 98

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : SPTREMBL21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	834	80.3	1039	16 Q9ANV1	Q9anv1 streptococc
2	32	3.1	802	16 Q97QM8	Q97qm8 streptococc
3	32	3.1	816	2 Q9AHY9	Q9ahy9 streptococc
4	32	3.1	819	2 Q9ANY3	Q9any3 streptococc
5	32	3.1	819	16 Q97QM9	Q97qm9 streptococc
6	32	3.1	839	16 Q9ANY2	Q9any2 streptococc
7	32	3.1	844	2 Q9AG74	Q9ag74 streptococc
8	16	1.5	289	2 Q9AE21	Q9ae21 streptococc
9	16	1.5	822	2 Q9ZHG7	Q9zhg7 streptococc
10	16	1.5	825	2 Q9ZGT5	Q9zgt5 streptococc
11	16	1.5	825	16 Q99XV4	Q99xv4 streptococc
12	9	0.9	198	5 Q8SSL6	Q8ssl6 encephalito
13	9	0.9	336	10 Q9LIM0	Q9lim0 arabidopsis
14	8	0.8	40	8 Q8WIJ2	Q8wij2 bostrychia
15	8	0.8	46	8 Q8WIJ0	Q8wij0 bostrychia
16	8	0.8	47	8 Q8WIJ8	Q8wij8 bostrychia

8	0.8	47	8	Q8WII9	bostrychia
8	0.8	47	8	Q8WII6	bostrychia
8	0.8	50	8	Q8WII6	bostrychia
8	0.8	50	8	Q8WII4	bostrychia
8	0.8	50	8	Q8WII3	bostrychia
8	0.8	50	8	Q8WII1	bostrychia
8	0.8	50	8	Q8WII8	bostrychia
8	0.8	50	8	Q8W929	bostrychia
8	0.8	50	8	Q8W823	bostrychia
8	0.8	50	8	Q8W822	bostrychia
8	0.8	50	8	Q8W8W8	bostrychia
8	0.8	50	8	Q8W8W7	bostrychia
8	0.8	131	5	Q9SVU3	Q9svu3 homarus ame
8	0.8	138	10	Q8S9N2	Q8s9n2 oryza sativ
8	0.8	139	2	Q65991	Q65991 clostridium
8	0.8	139	10	Q40683	Q40683 oryza sativ
8	0.8	139	10	Q22382	Q22382 oryza sativ
8	0.8	144	10	Q9ZTS2	Q9zts2 capsicum an
8	0.8	144	10	Q93XJ9	Q93xj9 solanum tub
8	0.8	146	16	Q97MN7	Q97mn7 clostridium
8	0.8	146	16	Q8RCS1	Q8rcs1 thermoanaer
8	0.8	154	16	Q8YEU1	Q8yeu1 bruceella me
8	0.8	182	8	Q8GFJ0	Q8gfj0 ginkgo bilo
8	0.8	185	8	Q8GFJ9	Q8gfj9 illicium pa
8	0.8	190	8	Q8GFJ2	Q8gfj2 dioscorea b
8	0.8	190	8	Q8GFJ6	Q8gfj6 saururus ce
8	0.8	199	8	Q8GFJ5	Q8gfj5 calycanthus
8	0.8	200	8	Q8GFJ6	Q8gfj6 cabomba car
8	0.8	201	8	Q8GFJ8	Q8gfj8 acorus cala
8	0.8	201	8	Q8GFJ7	Q8gfj7 asarum cana
8	0.8	201	8	Q8GFJ4	Q8gfj4 ceratophyll
8	0.8	201	8	Q8GFJ1	Q8gfj1 orimys wint
8	0.8	201	8	Q8GFJ8	Q8gfj8 lactoris fe
8	0.8	201	8	Q8GFJ7	Q8gfj7 liriodendro
8	0.8	201	8	Q9MSQ5	Q9msq5 zamia furfu
8	0.8	201	8	Q9MSP3	Q9msp3 nymphaea od
8	0.8	201	8	Q8GE23	Q8ge23 amborella t
8	0.8	207	17	Q29465	Q29465 archaeoglob
8	0.8	252	15	Q9IU18	Q9iu18 human immun
8	0.8	252	15	Q9IU03	Q9iu03 human immun
8	0.8	274	16	Q9RKB3	Q9rkb3 streptococ
8	0.8	278	16	Q8YBA6	Q8yba6 bruceella me
8	0.8	278	16	Q8UA71	Q8ua71 agrobacteri
8	0.8	312	10	Q94HD4	Q94hd4 oryza sativ
8	0.8	314	12	Q91MW4	Q91mw4 lumpy skin
8	0.8	327	2	Q9X303	Q9x303 bacillus an
8	0.8	345	16	Q9HX31	Q9hx31 pseudomonas
8	0.8	372	5	Q9U1I8	Q9u1i8 drosophila
8	0.8	410	16	Q8YU70	Q8yut0 anabaena sp
8	0.8	412	17	Q96XE9	Q96xe9 sulfolobus
8	0.8	472	2	Q46120	Q46120 campylobact
8	0.8	472	16	Q9PM69	Q9pm69 campylobact
8	0.8	514	2	Q9XSR9	Q9xsr9 streptomyce
8	0.8	522	16	Q9KF15	Q9kfl5 bacillus ha
8	0.8	543	4	Q9BY87	Q9by87 homo sapien
8	0.8	564	2	Q9R9P9	Q9r9p9 rhizobium m
8	0.8	573	16	Q97DG6	Q97dg6 clostridium
8	0.8	661	11	Q9JKS3	Q9jks3 mus musculu
8	0.8	672	3	Q66339	Q66339 saccharomyc
8	0.8	681	4	Q9UP10	Q9up10 homo sapien
8	0.8	695	10	Q8WSP0	Q8wsp0 oryza sativ
8	0.8	723	11	Q9VWH2	Q9vwh2 mus musculu
8	0.8	723	11	Q9JKS4	Q9jks4 mus musculu
8	0.8	743	5	Q23859	Q23859 dictyosteli
8	0.8	763	5	Q23874	Q23874 dictyosteli
8	0.8	765	3	Q94672	Q94672 schizosacch
8	0.8	778	5	Q8TIF5	Q8tif5 dictyosteli
8	0.8	791	16	Q9CL01	Q9cl01 pasteurella
8	0.8	793	10	Q9ASU6	Q9asu6 arabidopsis
8	0.8	813	10	Q9SHN6	Q9shn6 arabidopsis
8	0.8	840	15	Q9RSJ4	Q9rsj4 deinococcus
8	0.8	984	10	Q9LQY4	Q9lqy4 arabidopsis
8	0.8	1100	5	Q9VUE7	Q9vue7 drosophila

90 8 0.8 1148 2 Q60045 Q60045 thermoanaer
91 8 0.8 1210 11 Q91VW5 Q91VW5 mus musculus
92 8 0.8 1237 4 Q00114 Q00114 homo sapien
93 8 0.8 1308 4 Q60307 Q60307 homo sapien
94 8 0.8 1353 3 Q8X1F2 Q8X1F2 venturia in
95 8 0.8 1566 11 Q9R1L5 Q9R1L5 mus musculus
96 8 0.8 1583 4 Q9Y2H9 Q9Y2H9 homo sapien
97 8 0.8 2238 11 Q70365 Q70365 mus musculus
98 8 0.8 2986 12 Q91BL4 Q91BL4 perina nuda

ALIGNMENTS

RESULT 1
Q9ANV1 PRELIMINARY; PRT: 1039 AA.
AC Q9ANV1
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
DE Pneumococcal histidine triad protein E precursor (Hypothetical protein
DE SP1004)
GN PTE OR SP1004.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.,
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis";
RL Infect. Immun. 69:949-958(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=TIGR4;
RC MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Mayhew L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Mayhew L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Otterback T.R., Hansen C.L.,
RA Holtzapple E., Khouri H., Wolf A.M., Otterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001).
DR ENBL: AF318956; AAK06761.1;
DR ENBL: AF318956; AAK06761.1;
DR TIGR: SP1004;
FW Signal; Hypothetical protein; Complete proteome.
FT SIGNAL 1 POTENTIAL 29
SQ SEQUENCE 1039 AA; 114631 MW; 81A563FC806625C4 CRC64;

Query Match 80.3%; Score 834; DB 16; Length 1039;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1034; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKFSKVIAGSAVIVSLSCAVALNHRQENKNNRVSYVDGSSQSKSENITPDQVS 60
DB 1 MKFSKVIAGSAVIVSLSCAVALNHRQENKNNRVSYVDGSSQSKSENITPDQVS 60
QY 61 QKEGIAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDVALFSELLMKDPNQLKADIVN 120
DB 61 QKEGIAEQIVIKITDQGYVTSBGHDHYHYNGKVPYDVALFSELLMKDPNQLKADIVN 120
QY 121 EVKGGYIIKVGKYYVYLKDAHADNVRTKDEINRQKQEHVKNKNSNVAVARSQRY 180
DB 121 EVKGGYIIKVGKYYVYLKDAHADNVRTKDEINRQKQEHVKNKNSNVAVARSQRY 180

DB 121 EVKGGYIIKVGKYYVYLKDAHADNVRTKDEINRQKQEHVKNKNSNVAVARSQRY 180
QY 181 TTNDGYVFNPAIDITDTCNAYIVPHGGHYHYIPKSDLSASASAAKAHLAKGNQPSOLS 240
DB 181 TTNDGYVFNPAIDITDTCNAYIVPHGGHYHYIPKSDLSASASAAKAHLAKGNQPSOLS 240
QY 241 YSSTASDNTQSAVAGSTSKPANKSENQLSKLKYDPSAQRYSQSDGLVDFPAKIIIS 300
DB 241 YSSTASDNTQSAVAGSTSKPANKSENQLSKLKYDPSAQRYSQSDGLVDFPAKIIIS 300
QY 301 TPNGVAIPHGDHYHFIPIYSKLSALEETARMVPIISGTGTVSTNAKNEVVSSLGSSLN 360
DB 301 TPNGVAIPHGDHYHFIPIYSKLSALEETARMVPIISGTGTVSTNAKNEVVSSLGSSLN 360
QY 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLPNNSLA 420
DB 361 PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLPNNSLA 420
QY 421 TPSPSLP-IPGTSHEKHEEDGYGFDANRIIAEDSESGFVMSHGDHNIHFFKDLTEEQIKA 480
DB 421 TPSPSLP-IPGTSHEKHEEDGYGFDANRIIAEDSESGFVMSHGDHNIHFFKDLTEEQIKA 480
QY 481 AQKHEEVKTSNGSLSSHEQDYPGNAKEMKDLKKIEEKIAGIMQYGVKRESIVN 540
DB 481 AQKHEEVKTSNGSLSSHEQDYPGNAKEMKDLKKIEEKIAGIMQYGVKRESIVN 540
QY 541 KEKNAIYPHGDHHDHAPIDEHKPVGIGHSHSNVELFKPEEGVAKKGNKYVTEELTNV 600
DB 541 KEKNAIYPHGDHHDHAPIDEHKPVGIGHSHSNVELFKPEEGVAKKGNKYVTEELTNV 600
QY 601 VNLKNSFNQNF-FLANGQKRVSPFPPELEKLGINLVKLITPDGKVLKYGKVF 660
DB 601 VNLKNSFNQNF-FLANGQKRVSPFPPELEKLGINLVKLITPDGKVLKYGKVF 660
QY 661 EGVGNIAFELDQPLPGQTKYTIASKDYPEVSDGTFTVPTSLAYKMSQTIYFPFHA 720
DB 661 EGVGNIAFELDQPLPGQTKYTIASKDYPEVSDGTFTVPTSLAYKMSQTIYFPFHA 720
QY 721 GTYLRVNPQFAVPKGTALYRVDFEFGNAYLENNYKVGKIKLPKLNQGTTRTACNK 780
DB 721 GTYLRVNPQFAVPKGTALYRVDFEFGNAYLENNYKVGKIKLPKLNQGTTRTACNK 780
QY 781 IPTVTANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRKAQENSKLDEKVEEPKTS 840
DB 781 IPTVTANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRKAQENSKLDEKVEEPKTS 840
QY 841 ERVEKEKISGTSNSTSLTLEVPVDPVQKVAFAESYGNKLENVLFNMDGTIELYLP 900
DB 841 ERVEKEKISGTSNSTSLTLEVPVDPVQKVAFAESYGNKLENVLFNMDGTIELYLP 900
QY 901 SGEVILKKNMADFTGEAPOGNGENKPSSENGKSTGTVENQPTENKPADSLPEAPNEKPVKP 960
DB 901 SGEVILKKNMADFTGEAPOGNGENKPSSENGKSTGTVENQPTENKPADSLPEAPNEKPVKP 960
QY 961 ENSTDNGMLNPEGVSGDPMLEAPADVPDVEKLEKFTASYGLDSDVIFNMDGTI 1020
DB 961 ENSTDNGMLNPEGVSGDPMLEAPADVPDVEKLEKFTASYGLDSDVIFNMDGTI 1020
QY 1021 EURLPSGEVILKKNLSD 1036
DB 1021 EURLPSGEVILKKNLSD 1036
RESULT 2
Q97OM8 PRELIMINARY; PRT: 802 AA.
AC Q97OM8
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Conserved domain protein.
GN SP1175.
OS Streptococcus pneumoniae.

DT	01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE	01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DD	Pneumococcal histidine triad protein B precursor (Fragment).
GN	PHB.
OS	Streptococcus pneumoniae.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC	Streptococcaceae; Streptococcus.
OX	NCBI_TaxID=1313;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=21101045; PubMed=11159990;
RA	Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA	Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA	Langermann S., Koenig S., Johnson S.;
RT	"Identification and characterization of a novel family of pneumococcal
RT	proteins (the pnt family) that are protective against sepsis.";
RL	Infect. Immun. 69:949-958(2001).
DR	EMBL: AF318954; AAK06759.1; -
KW	Signal.
FT	SIGNAL
FT	NON_TER 819 819
SQ	SEQUENCE 819 AA; 92108 MW; E602CFC16CC28A5F CRC64;
Query Match	3.1%; Score 32; DB 2; Length 819;
Best Local Similarity	100.0%; Pred. No. 2.4e+23;
Matches	32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	67 AEQIVKITDQGYVTSKGDHYHNGKVPYDA 98
Db	67 AEQIVKITDQGYVTSKGDHYHNGKVPYDA 98
RESULT 5	
Q97QM9	
ID	PRELIMINARY; PRT; 819*AA.
AC	Q97QM9;
DT	01-OCT-2001 (TReMBLrel. 18, Created)
DD	01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE	Conserved domain protein.
GN	SP1174.
OS	Streptococcus pneumoniae.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC	Streptococcaceae; Streptococcus.
OX	NCBI_TaxID=1313;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=TIGR4;
RX	MEDLINE=11357209; PubMed=11463916;
RA	Tattelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA	Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA	Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA	Mayway L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA	Holtzapfe E., Khouri H., Wolf A.M., Uitterback T.R., Hansen C.L.,
RA	McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT	"Complete genome sequence of a virulent isolate of Streptococcus
RT	pneumoniae.";
RL	Science 293:498-506(2001);
DR	EMBL: AE007418; AAK75283.1; -
DR	TIGR: SP1174; -
KW	Complete proteome.
SQ	SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;
Query Match	3.1%; Score 32; DB 16; Length 819;
Best Local Similarity	100.0%; Pred. No. 2.4e+23;
Matches	32; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY	67 AEQIVKITDQGYVTSKGDHYHNGKVPYDA 98
Db	67 AEQIVKITDQGYVTSKGDHYHNGKVPYDA 98

RA Green B.A.;
RT "Recombinant PtpA protein, a Unique Histidine Motif-Containing Protein
from Streptococcus pneumoniae, Protects Mice against Intranasal
Pneumococcal Challenge";
RL Infect. Immun. 69:3827-3836(2001).
DR EMBL: AF340221; AAK26629.1; -
SQ SEQUENCE 844 AA; 94769 MW; D738A55290FF8902 CRC64;
Query Match 3.1%; Score 32; DB 2; Length 844;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 67 AEQIVIKITDQGYVTSFGDHYHYNGKVPYDA 98
DB 66 AEQIVIKITDQGYVTSFGDHYHYNGKVPYDA 97
RESULT 8
Q9AE21 PRELIMINARY; PRT; 289 AA.
AC Q9AE21; 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Hypothetical 32.0 kDa protein (Fragment).
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5531;
RX MEDLINE=21172873; PubMed=11274116;
RA Granlund M., Michel F., Norrgren M.;
RT "Mutually exclusive distribution of IS1548 and GBS1, an active group
II intron identified in human isolates of group B streptococci";
RL J. Bacteriol. 183:2560-2569(2001).
DR EMBL: AJ290952; CAC35985.1; -
KW Hypothetical protein.
FT NON_TER 289
SQ SEQUENCE 289 AA; 32043 MW; A15A8588EA8140E4 CRC64;
Query Match 1.5%; Score 16; DB 2; Length 289;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 73 KITDQGYVTSFGDHYH 88
DB 75 KITDQGYVTSFGDHYH 90
RESULT 9
Q9ZHG7 PRELIMINARY; PRT; 822 AA.
AC Q9ZHG7;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical 92.4 kDa protein.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R268;
RX MEDLINE=99113568; PubMed=9916102;
RA Spellerberg B., Rozdzinski E., Martin S., Weber-Heymann J.,
RA Schnitzler N., Luetticken R., Podbielski A.;
RT "Lmb, a protein with similarities to the Lrai adhesin family, mediates
attachment of Streptococcus agalactiae to human laminin";
RL Infect. Immun. 67:871-878(1999).
DR EMBL: AF062533; AAD13797.1; -

RA Q9ANY2 PRELIMINARY; PRT; 839 AA.
AC Q9ANY2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE pneumococcal histidine triad protein D precursor (Hypothetical protein
SPI003) (Fragment).
GN PHTD OR SPI003.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
proteins (the Pht family) that are protective against sepsis";
RL Infect. Immun. 69:949-958(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae";
RL Science 293:498-506(2001).
DR EMBL: AF318955; AAK06760.1; -
DR EMBL: AE007403; AAK75120.1; -
KW Signal; Hypothetical protein; Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 839 839
SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;
Query Match 3.1%; Score 32; DB 16; Length 839;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 67 AEQIVIKITDQGYVTSFGDHYHYNGKVPYDA 98
DB 67 AEQIVIKITDQGYVTSFGDHYHYNGKVPYDA 98
RESULT 7
Q9AG74 PRELIMINARY; PRT; 844 AA.
AC Q9AG74;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PtpA.
GN PtpA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21246685; PubMed=11349048;
RA Zhang Y., Masi A.W., Barniak V., Mountzouras K., Hostetter M.K.,

```

KW Hypothetical protein.
SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;

Query Match 1.5%; Score 16; DB 2; Length 822;
Best Local Similarity 100.0%; Pred. No. 6.4e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 KITDQGYVTSHGDHYH 88
DB 75 KITDQGYVTSHGDHYH 90

RESULT 10
Q93GT5 PRELIMINARY; PRT; 825 AA.
AC Q93GT5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Histidine triad protein of group A streptococci.
GN HTPA.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-9;
RA Terao Y., Kawabata S., Hamada S.;
RT "Characterization of a novel histidine triad protein of group A
RT streptococci."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB073855; BAB1774.1; -
SQ SEQUENCE 825 AA; 92623 MW; DE4ECC199181DFFB CRC64;

Query Match 1.5%; Score 16; DB 2; Length 825;
Best Local Similarity 100.0%; Pred. No. 6.4e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 KITDQGYVTSHGDHYH 88
DB 75 KITDQGYVTSHGDHYH 90

RESULT 11
Q99XV4 PRELIMINARY; PRT; 825 AA.
AC Q99XV4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein SPV2006.
GN SPV2006.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RA MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4638-4663(2001).
DR EMBL: AE006623; AAK34688.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 825 AA; 92649 MW; 57DF50969E6D50F4 CRC64;

Query Match 1.5%; Score 16; DB 16; Length 825;
Best Local Similarity 100.0%; Pred. No. 6.4e-07;

```

```

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 KITDQGYVTSHGDHYH 88
DB 75 KITDQGYVTSHGDHYH 90

RESULT 12
Q8SS16 PRELIMINARY; PRT; 198 AA.
AC Q8SS16;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 60S ribosomal protein L13A (L16).
GN EC004.1380.
OS Encephalitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Prentier G., Barbe V., Peyretallade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi."
RL Nature 414:450-453(2001).
DR EMBL: AL590444; CAD25327.1; -
SQ SEQUENCE 198 AA; 22679 MW; 267BC81D1C78FE2B CRC64;

Query Match 0.9%; Score 9; DB 5; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 KDLKKTIEE 521
DB 171 KDLKKTIEE 179

RESULT 13
Q9LIM0 PRELIMINARY; PRT; 336 AA.
AC Q9LIM0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Emb|CAB09999.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones."

```

```

RL  DNA Res. 7:217-221(2000).
DR  EMBL; AF382906; AAL67234.1;
SQ  SEQUENCE 336 AA: 37575 MW; 4BCAC3731EB66F43 CRC64;

Query Match      0.8%; Score 9; DB 10; Length 336;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  804 LEXENOTDK 812
DB  133 LEXENOTDK 141
|||||

RESULT 14
Q8WIJ2 PRELIMINARY; PRT; 40 AA.
AC Q8WIJ2;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P3880.QLD;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
(Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
mitochondrial and plastid DNA markers."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382902; AAL67226.1;
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER 40
SQ SEQUENCE 40 AA: 4723 MW; 96013F36C758F617 CRC64;

Query Match      0.8%; Score 8; DB 8; Length 40;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  472 DLTEEQIK 479
DB  13 DLTEEQIK 20
|||||

RESULT 15
Q8WIJ0 PRELIMINARY; PRT; 46 AA.
AC Q8WIJ0;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P2962.SG;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
(Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
mitochondrial and plastid DNA markers."

```

```

RL  Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AF382906; AAL67234.1;
DR  InterPro; IPR000894; RuBisCO_small.
DR  Pfam; PF00101; RuBisCO_small; 1.
DR  ProDom; PD000290; RuBisCO_small; 1.
KW  Chloroplast.
FT  NON_TER 46
SQ  SEQUENCE 46 AA: 5345 MW; 6E8EDA7D7D489216 CRC64;

Query Match      0.8%; Score 8; DB 8; Length 46;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  472 DLTEEQIK 479
DB  13 DLTEEQIK 20
|||||

RESULT 16
Q8WIJ8 PRELIMINARY; PRT; 47 AA.
AC Q8WIJ8;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C.E349.PA;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
(Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
mitochondrial and plastid DNA markers."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382893; AAL67208.1;
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER 47
SQ SEQUENCE 47 AA: 5575 MW; 65F19EFC0529892 CRC64;

Query Match      0.8%; Score 8; DB 8; Length 47;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  472 DLTEEQIK 479
DB  13 DLTEEQIK 20
|||||

RESULT 17
Q8WIJ9 PRELIMINARY; PRT; 47 AA.
AC Q8WIJ9;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=P4097.MW;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodmelaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382907; AAL67236.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 47 AA; 5508 MW; 65F8ED47D7D4892 CRC64;

Query Match 0.8%; Score 8; DB 8; Length 47;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 472 DLTEEQIK 479
Db 13 DLTEEQIK 20

RESULT 18
Q8WIJ6
ID Q8WIJ6 PRELIMINARY; PRT; 47 AA.
AC Q8WIJ6;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodmelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3054.BZ;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodmelaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382909; AAL67240.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 47 AA; 5573 MW; 8B3C91CFEC7643AA CRC64;

Query Match 0.8%; Score 8; DB 8; Length 47;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 472 DLTEEQIK 479
Db 13 DLTEEQIK 20

RESULT 19
Q8WIJ6
ID Q8WIJ6 PRELIMINARY; PRT; 50 AA.
AC Q8WIJ6;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.

```

```

OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodmelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodmelaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382897; AAL67216.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 5985 MW; 1394B315F18E2680 CRC64;

Query Match 0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 472 DLTEEQIK 479
Db 13 DLTEEQIK 20

RESULT 20
Q8WIJ4
ID Q8WIJ4 PRELIMINARY; PRT; 50 AA.
AC Q8WIJ4;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodmelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodmelaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382898; AAL67218.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 6001 MW; 655E3D5D5F18E2690 CRC64;

Query Match 0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 472 DLTEEQIK 479
Db 13 DLTEEQIK 20

RESULT 21
Q8WIJ3
ID Q8WIJ3 PRELIMINARY; PRT; 50 AA.
AC Q8WIJ3;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)

```

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBGS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P3065.BZ;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382899; AAL67220.1; -
DR InterPro: IPR000894; RuBisCO_small.
DR Pfam: PF00101; RuBisCO_small; 1.
DR ProDom: PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 6004 MW; 13854315F18E2680 CRC64;

Query Match 0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20
|||||
RESULT 22
Q8WJ1 PRELIMINARY; PRT; 50 AA.
AC Q8WJ1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBGS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P2961.SG;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382905; AAL67232.1; -
DR InterPro: IPR000894; RuBisCO_small.
DR Pfam: PF00101; RuBisCO_small; 1.
DR ProDom: PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 5037 MW; 2CBA4315FE8ED47D CRC64;

Query Match 0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20
|||||
RESULT 23

Q8WJ18 PRELIMINARY; PRT; 50 AA.
AC Q8WJ18;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBGS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P3065.BZ;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382908; AAL67238.1; -
DR InterPro: IPR000894; RuBisCO_small.
DR Pfam: PF00101; RuBisCO_small; 1.
DR ProDom: PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 5948 MW; 23E5337B222F610C CRC64;

Query Match 0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20
|||||
RESULT 24
Q8W929 PRELIMINARY; PRT; 50 AA.
AC Q8W929;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBGS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3190.CO, C3194.CO, C3252.MX, P4122.FL, C3400.MX, AND C3191.CO;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF382890; AAL67202.1; -
DR EMBL; AF382891; AAL67204.1; -
DR EMBL; AF382892; AAL67206.1; -
DR EMBL; AF382894; AAL67210.1; -
DR EMBL; AF382895; AAL67212.1; -
DR EMBL; AF382896; AAL67214.1; -
DR InterPro: IPR000894; RuBisCO_small.
DR Pfam: PF00101; RuBisCO_small; 1.
DR ProDom: PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 6004 MW; 13854315F19EFCE0 CRC64;


```

Query Match          0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20
|||||||

RESULT 25
ID Q8W823 PRELIMINARY; PRT; 50 AA.
AC Q8W823;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P3648.BZ, C3369.GU, P3016.PE, AND P3414.FL;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
(Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF382910; AAL67242.1; -.
DR EMBL; AF382911; AAL67244.1; -.
DR EMBL; AF382912; AAL67246.1; -.
DR EMBL; AF382913; AAL67248.1; -.
DR Toterpro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 6002 MW; 23E5337B3C91CFEC CRC64;

Query Match          0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20
|||||||

RESULT 26
ID Q8W822 PRELIMINARY; PRT; 50 AA.
AC Q8W822;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P3216.OLD, AND P3999.NT;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
(Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.

```

```

DR EMBL; AF382915; AAL67252.1; -.
DR EMBL; AF382916; AAL67254.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 6060 MW; 23E5337B3EFC5FEC CRC64;

Query Match          0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20
|||||||

RESULT 27
ID Q8W8W8 PRELIMINARY; PRT; 50 AA.
AC Q8W8W8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P2812.OLD, AND P2817.OLD;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
(Rhodomelaceae, Rhodophyta) and divergence rates based on nuclear,
mitochondrial and plastid DNA markers.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF382900; AAL67222.1; -.
DR EMBL; AF382901; AAL67224.1; -.
DR InterPro; IPR000894; RuBisCO_small.
DR Pfam; PF00101; RuBisCO_small; 1.
DR ProDom; PD000290; RuBisCO_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA; 6009 MW; 2CBA430B0D0C447D CRC64;

Query Match          0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEEQIK 479
Db 13 DLTEEQIK 20
|||||||

RESULT 28
ID Q8W8W7 PRELIMINARY; PRT; 50 AA.
AC Q8W8W7;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Ribulose-1,5-bisphosphate small subunit (Fragment).
GN RBCS.
OS Bostrychia calliptera.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Florideophyceae; Ceramiales; Rhodomelaceae;
OC Bostrychia.
OX NCBI_TaxID=161377;
RN [1]

```

RN SEQUENCE FROM N.A.
RC STRAIN-P128.NT, AND P3752.NT;
RA Zuccarello G.C., West J.A.;
RT "Phylogeography of the Bostrychia calliptera/B. pinnata complex
RT (Rhodomeleaceae, Rhodophyta) and divergence rates based on nuclear,
RT mitochondrial and plastid DNA markers."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF382903; AAL67228.1;
DR EMBL: AF382904; AAL67230.1;
DR InterPro: IPR000894; Rubisco_small.
DR Pfam: PF00101; Rubisco_small; 1.
DR ProDom: PD000290; Rubisco_small; 1.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 50 AA: 5995 MW: 2CBA4315FCE3447D CRC64;

Query Match 0.8%; Score 8; DB 8; Length 50;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEQIK 479
Db 13 DLTEQIK 20

RESULT 29
Q95VU3 PRELIMINARY; PRT: 131 AA.
AC Q95VU3;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Olfactory enriched transcript 09.47 (Fragment).
OS Homarus americanus (American lobster).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Nephropoidea; Nephropidae; Homarus.
OX NCBI_TaxID=6706;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-OLFACTORY ORGAN;
RA McClintock T.S., Schweder D., Hollins B.;
RT "Olfactory enriched transcripts are cell type specific markers in the
RT lobster olfactory organ."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF405437; AAL04104.1;
FT NON_TER 1 131
FT NON_TER 131 131
SQ SEQUENCE 131 AA: 15093 MW: D9F37C14BDAL2A9F CRC64;

Query Match 0.8%; Score 8; DB 5; Length 131;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 LKDADIVN 120
Db 111 LKDADIVN 118

RESULT 30
Q8S9N2 PRELIMINARY; PRT: 138 AA.
AC Q8S9N2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Anti-disease protein 1.
GN ADII.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;

RN SEQUENCE FROM N.A.
RP Rihe P., Qunhong Y., Aisheng X., Xian L., Huiguo F.;
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY072818; AAL77198.1;
SQ SEQUENCE 138 AA: 14864 MW: 0AL84EE33BCEB94A CRC64;
Query Match 0.8%; Score 8; DB 10; Length 138;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 641 VKLITPDG 648
Db 47 VKLITPDG 54

RESULT 31
O65991 PRELIMINARY; PRT: 139 AA.
AC O65991;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MTLF.
GN MTLF.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 792;
RX MEDLINE=21097245; PubMed=11160802;
RA Behrens S., Mitchell W.J., Bahl H.;
RT "Molecular analysis of the mannitol operon of Clostridium
RT acetobutylicum encoding a phosphotransferase system and a putative
RT PTS-modulated regulator."
RL Microbiology 147:75-86(2001).
DR EMBL: U53868; AAC12850.1;
DR HSSP: P00550; 1A3A.
DR InterPro: IPR002178; PTS_EIIA_2.
DR Pfam: PF00359; PTS_EIIA_2; 1.
DR ProDom: PD001689; PTS_EIIA_2; 1.
DR ProSITE: PS00372; PTS_EIIA_2; 1.
SQ SEQUENCE 139 AA: 15676 MW: 09471CAC67903B11 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
Db 58 NGVAIPHG 65

RESULT 32
Q40683 PRELIMINARY; PRT: 139 AA.
AC Q40683;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Ferredoxin.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KINMAZE; TISSUE=LEAF;
RA Ohmori K., Doyama N., Ida S.;
RT "Molecular cloning of a rice leaf ferredoxin cDNA.";

```

RL Plant Physiol. 111:348-348(1996).
CC -1- COFACTOR: BINDS A 2FE-2S CLUSTER (BY SIMILARITY).
DR EMBL: D30763; BAA06436.1; -.
DR HSSP: P00221; 1A70.
DR InterPro: IPR000564; 2Fe2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SFRDXIN.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Iron-sulfur.
SQ SEQUENCE 139 AA; 14951 MW; 301F47841265E93C CRC64;

Query Match 0.8%; Score 8; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 48 VKLITPDG 55

RESULT 33
O22382 PRELIMINARY; PRT; 139 AA.
AC O22382
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Ferredoxin.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MILYANG 23; TISSUE=IMMATURE SEED;
RA Lee M.C., Kim C.S., Yi B.Y., Eun M.Y.;
RT "Molecular cloning and characterization of ferredoxin gene from
RT rice."
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: BINDS A 2FE-2S CLUSTER (BY SIMILARITY).
DR EMBL: AF010320; AAB65699.1; -.
DR HSSP: P00221; 1A70.
DR InterPro: IPR000564; 2Fe2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SFRDXIN.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Iron-sulfur.
SQ SEQUENCE 139 AA; 14945 MW; 4236DF708E02C0D2 CRC64;

Query Match 0.8%; Score 8; DB 10; Length 139;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 48 VKLITPDG 55

RESULT 34
Q9ZTS2 PRELIMINARY; PRT; 144 AA.
AC Q9ZTS2
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Ferredoxin-like protein.
GN AP1.
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

```

OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECW;
RA Lin H.J., Lee B.H., Lee Y.S., Pai C.H., Lin N.S., Feng T.Y.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: BINDS A 2FE-2S CLUSTER (BY SIMILARITY).
DR EMBL: AF039562; AAD02175.1; -.
DR HSSP: P00221; 1A70.
DR InterPro: IPR000564; 2Fe2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PRINTS: PR00159; 2FE2SFRDXIN.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
KW Iron-sulfur.
SQ SEQUENCE 144 AA; 15178 MW; E9763663281D58EB CRC64;

Query Match 0.8%; Score 8; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 52 VKLITPDG 59

RESULT 35
Q93XJ9 PRELIMINARY; PRT; 144 AA.
AC Q93XJ9
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Ferredoxin I.
GN FED1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Holtgreife S., Bader K.P., Horton P., Scheibe R., von Schaewen A.,
RA Backhausen J.E.;
RT "Alterations of leaf ferredoxin (Fd I) contents in transgenic potato
RT (Solanum tuberosum L.) plants changes electron distribution, and
RT influences light acclimation of transgenic plants."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ307031; CAC38395.1; -.
DR InterPro: IPR000564; 2Fe2S_ferredoxin.
DR InterPro: IPR001041; Ferredoxin.
DR Pfam: PF00111; fer2; 1.
DR PROSITE: PS00197; 2FE2S_FERREDOXIN; UNKNOWN_1.
KW Iron-sulfur.
FT CHAIN 48 144 FERREDOXIN I.
SQ SEQUENCE 144 AA; 15363 MW; C06025EA1B257128 CRC64;

Query Match 0.8%; Score 8; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 52 VKLITPDG 59

RESULT 36
Q97MN7 PRELIMINARY; PRT; 146 AA.
AC Q97MN7
DT 01-OCT-2001 (TRENBLrel. 18, Created)

```

```

DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
PTS system, mannitol-specific IIA domain (Ntr-type) (gene
DE MTF).
DE CAC0156.
GN Clostridium acetobutylicum.
OS Clostridia: Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Koonin E.V., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AF007528; AAK78140.1; -.
DR InterPro: IPR002178; PTS_EIIA_2.
DR Pfam: PF00359; PTS_EIIA_2; 1.
DR ProDom: PD001689; PTS_EIIA_2; 1.
DR PROSITE: PS00372; PTS_EIIA_2; 1.
KW Complete proteome.
SQ SEQUENCE 146 AA; 16318 MW; A9A1C73CCD74DF97 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 146;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
DB 58 NGVAIPHG 55
|||||

RESULT 37
Q8RCS1 PRELIMINARY; PRT; 146 AA.
AC Q8RCS1;
DT 01-JUN-2002 (TREMELrel. 21, Created)
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Phosphotransferase system mannitol/fructose-specific IIA domain
DE (Ntr-type).
GN PTSN2 OR TTE0341.
OS Thermoanaerobacter tengcongensis.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
DR EMBL: AF013006; AAM23633.1; -.
KW Transferase; Complete proteome.
SQ SEQUENCE 146 AA; 16368 MW; ADB28A45BEB8835 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 146;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
DB 58 NGVAIPHG 55
|||||

```

```

RESULT 38
Q8YEUI PRELIMINARY; PRT; 154 AA.
AC Q8YEUI;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Nitrogen regulatory IIA protein (EC 2.7.1.69).
GN BME11786.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / B10TYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapratral V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.J.,
RA Haselkorn R., Kyripides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL: AE009611; AAL52967.1; -.
DR InterPro: IPR004715; PTSIIA fruc.
DR InterPro: IPR002178; PTS_EIIA_2.
DR Pfam: PF00359; PTS_EIIA_2; 1.
DR ProDom: PD001689; PTS_EIIA_2; 1.
DR TIGRFAMS: TIGR00848; fruc. 1.
DR PROSITE: PS00372; PTS_EIIA_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 154 AA; 16497 MW; D53523086D684FB4 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 154;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 303 NGVAIPHG 310
DB 60 NGVAIPHG 67
|||||

RESULT 39
Q9GFJO PRELIMINARY; PRT; 182 AA.
AC Q9GFJO;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Ribosomal protein L2 (fragment).
GN RPL2.
OS Ginkgo biloba (Ginkgo).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Ginkgoophyta; Ginkgoales; Ginkgoaceae; Ginkgo.
OX NCBI_TaxID=3311;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11080123;
RA Graham S.W., Olmstead R.G.;
RT "Utility of 17 chloroplast genes for inferring the phylogeny of the
basal angiosperms."
KL Am. J. Bot. 87:1712-1730(2000).
DR EMBL: AF123793; AAC26141.1; -.
DR HSP; P04257; IRL2.
DR InterPro: IPR002171; Ribosomal_L2.
DR Pfam: PF00181; Ribosomal_L2; 1.
DR TIGRFAMS: TIGR01171; rplB_bact; 1.
DR PROSITE: PS00467; RIBOSOMAL_L2; 1.
KW Chloroplast.
ET NON_TER 1 1

```

```

RT  NON_TER      182      182
SQ  SEQUENCE     182 AA: 19540 MW: 765F71F0F61CB89E CRC64;

Query Match          0.8%; Score 8; DB 8; Length 182;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1022 LRLPSGEV 1029
DB  115 LRLPSGEV 122

RESULT 40
Q9GFJ9
ID  Q9GFJ9; PRELIMINARY; PRT: 185 AA.
AC  Q9GFJ9;
DT  01-MAR-2001 (TREMBLrel. 16, Created)
DT  01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT  01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE  Ribosomal protein L2 (Fragment).
GN  RPL2.
OS  Illicium parviflorum.
OG  Chloroplast.
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Illiciales; Illiciaceae; Illicium.
OX  NCBI_TaxID=13099;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  PubMed=11080123;
RA  Graham S.W., Olmstead R.G.;
RT  "Utility of 17 chloroplast genes for inferring the phylogeny of the
RT  basal angiosperms.";
RL  Am. J. Bot. 87:1712-1730(2000).
DR  EMBL; AF123794; AAG26142.1; -.
DR  HSSP; P04257; 1RL2.
DR  InterPro; IPR002171; Ribosomal_L2.
DR  Pfam; PF00181; Ribosomal_L2; 1.
DR  TIGRFAMS; TIGS01171; rplB_bact; 1.
DR  PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW  Chloroplast.
FT  NON_TER      1      185
FT  NON_TER      185
SQ  SEQUENCE     185 AA: 19810 MW: D6D0ARCFBCE8E8630 CRC64;

Query Match          0.8%; Score 8; DB 8; Length 185;
Best Local Similarity 100.0%; Pred.No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1022 LRLPSGEV 1029
DB  118 LRLPSGEV 125

RESULT 41
Q9GFJ2
ID  Q9GFJ2; PRELIMINARY; PRT: 190 AA.
AC  Q9GFJ2;
DT  01-MAR-2001 (TREMBLrel. 16, Created)
DT  01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT  01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE  Ribosomal protein L2 (Fragment).
GN  RPL2.
OS  Dioscorea bulbifera (Air potato).
OG  Chloroplast.
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Liliopsida; Dioscoreales; Dioscoreaceae;
OX  Dioscorea.
OX  NCBI_TaxID=35874;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  PubMed=11080123;
RA  Graham S.W., Olmstead R.G.;
RT  "Utility of 17 chloroplast genes for inferring the phylogeny of the

```

```

RT  basal angiosperms.";
RL  Am. J. Bot. 87:1712-1730(2000).
DR  EMBL; AF123791; AAG26139.1; -.
DR  HSSP; P04257; 1RL2.
DR  InterPro; IPR002171; Ribosomal_L2.
DR  Pfam; PF00181; Ribosomal_L2; 1.
DR  TIGRFAMS; TIGR01171; rplB_bact; 1.
DR  PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW  Chloroplast.
FT  NON_TER      1      190
FT  NON_TER      190
SQ  SEQUENCE     190 AA: 20210 MW: D51F3555C9439798 CRC64;

Query Match          0.8%; Score 8; DB 8; Length 190;
Best Local Similarity 100.0%; Pred.No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1022 LRLPSGEV 1029
DB  124 LRLPSGEV 131

RESULT 42
Q9GFJ6
ID  Q9GFJ6; PRELIMINARY; PRT: 197 AA.
AC  Q9GFJ6;
DT  01-MAR-2001 (TREMBLrel. 16, Created)
DT  01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT  01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE  Ribosomal protein L2 (Fragment).
GN  RPL2.
OS  Saururus cernuus (Lizardtail).
OG  Chloroplast.
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Piperales; Saururaceae; Saururus.
OX  NCBI_TaxID=13260;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  PubMed=11080123;
RA  Graham S.W., Olmstead R.G.;
RT  "Utility of 17 chloroplast genes for inferring the phylogeny of the
RT  basal angiosperms.";
RL  Am. J. Bot. 87:1712-1730(2000).
DR  EMBL; AF123797; AAG26145.1; -.
DR  HSSP; P04257; 1RL2.
DR  InterPro; IPR002171; Ribosomal_L2.
DR  Pfam; PF00181; Ribosomal_L2; 1.
DR  TIGRFAMS; TIGR01171; rplB_bact; 1.
DR  PROSITE; PS00467; RIBOSOMAL_L2; 1.
KW  Chloroplast.
FT  NON_TER      1      197
FT  NON_TER      197
SQ  SEQUENCE     197 AA: 20965 MW: 77FE203A3504949B CRC64;

Query Match          0.8%; Score 8; DB 8; Length 197;
Best Local Similarity 100.0%; Pred.No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1022 LRLPSGEV 1029
DB  125 LRLPSGEV 132

RESULT 43
Q9GFJ5
ID  Q9GFJ5; PRELIMINARY; PRT: 199 AA.
AC  Q9GFJ5;
DT  01-MAR-2001 (TREMBLrel. 16, Created)
DT  01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT  01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE  Ribosomal protein L2 (Fragment).
GN  RPL2.
OS  Calycanthus floridus (Sweet shrub).

```

```

OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Nerales; Calycanthaceae; Calycanthus.
CX NCBI_TaxID=3429;
RN [1]
RP SEQUENCE FROM N.A.
RA Am. J. Bot. 87:1712-1730(2000).
RX PubMed=11080123;
RA Graham S.W., Olmstead R.G.;
RT "Utility of 17 chloroplast genes for inferring the phylogeny of the
RT basal angiosperms.";
RL Am. J. Bot. 87:1712-1730(2000).
DR EMBL; AF123788; AAG26136.1; -.
DR HSSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRfams; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00457; RIBOSOMAL_L2; 1.
KW Chloroplast.
FT NON_TER 1
FT NON_TER 199
SQ SEQUENCE 199 AA; 21316 MW; 2D2673BC6CF20DCD CRC64;

Query Match 0.8%; Score 8; DB 8; Length 199;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 124 LRLPSGEV 131

RESULT 44
Q9GFJ6 PRELIMINARY; PRT; 200 AA.
AC Q9GFJ6;
AT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ribosomal protein L2 (Fragment).
GN RPL2.
OS Cabomba caroliniana.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Cabombaceae; Cabomba.
CX NCBI_TaxID=4426;
RN [1]
RP SEQUENCE FROM N.A.
RA Am. J. Bot. 87:1712-1730(2000).
RX PubMed=11080123;
RA Graham S.W., Olmstead R.G.;
RT "Utility of 17 chloroplast genes for inferring the phylogeny of the
RT basal angiosperms.";
RL Am. J. Bot. 87:1712-1730(2000).
DR EMBL; AF123787; AAG26135.1; -.
DR HSSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRfams; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00457; RIBOSOMAL_L2; 1.
KW Chloroplast.
FT NON_TER 1
FT NON_TER 200
SQ SEQUENCE 200 AA; 21334 MW; 41F8BE657EF9D54 CRC64;

Query Match 0.8%; Score 8; DB 8; Length 200;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 126 LRLPSGEV 133

RESULT 45

```

```

Q9GFJ8 PRELIMINARY; PRT; 201 AA.
AC Q9GFJ8;
AT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ribosomal protein L2 (Fragment).
GN RPL2.
OS Acorus calamus (Sweet flag).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Acoraceae; Acorus.
CX NCBI_TaxID=4465;
RN [1]
RP SEQUENCE FROM N.A.
RA Am. J. Bot. 87:1712-1730(2000).
RX PubMed=11080123;
RA Graham S.W., Olmstead R.G.;
RT "Utility of 17 chloroplast genes for inferring the phylogeny of the
RT basal angiosperms.";
RL Am. J. Bot. 87:1712-1730(2000).
DR EMBL; AF123785; AAG26133.1; -.
DR HSSP; P04257; 1RL2.
DR InterPro; IPR002171; Ribosomal_L2.
DR Pfam; PF00181; Ribosomal_L2; 1.
DR TIGRfams; TIGR01171; rplB_bact; 1.
DR PROSITE; PS00457; RIBOSOMAL_L2; 1.
KW Chloroplast.
FT NON_TER 1
FT NON_TER 201
SQ SEQUENCE 201 AA; 21349 MW; 15441A79378F5459 CRC64;

Query Match 0.8%; Score 8; DB 8; Length 201;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1022 LRLPSGEV 1029
DB 126 LRLPSGEV 133

Search completed: May 13, 2003, 14:01:05
Job time : 55 secs

```

GenCore version 5.1.4_p5_4578
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:59:18 ; Search time 19 Seconds
(without alignments)
1608.969 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 1039

Sequence: 1 MKFSKKYIAGSAVIVSLSL.....IELRLPSGEVKKNLSDFIA 1039

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 8

Total number of hits satisfying chosen parameters: 5

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : Issued Patents.AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	447	43.0	447	4	US-08-961-083-182
2	32	3.1	796	4	US-08-961-083-56
3	28	2.7	763	4	US-08-961-083-66
4	8	0.8	144	2	US-09-049-577-2
5	8	0.8	144	4	US-09-390-598-2

ALIGNMENTS

RESULT 1
US-08-961-083-182
; Sequence 182, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-182

Query Match 43.0%; Score 447; DB 4; Length 447;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 LNHRSSENKNNRVSYVDGSSQSKSENLTPOVSOKEGIAEQIVIKITDQGYVTSHG 84

Db 1 LNHRSSENKNNRVSYVDGSSQSKSENLTPOVSOKEGIAEQIVIKITDQGYVTSHG 60

QY 85 DHYHYNGKVPYDALFSEELLKDPNYQLKDAIVNEVKGYYIKVDGKYVYVYLKDAHA 144

Db 61 DHYHYNGKVPYDALFSEELLKDPNYQLKDAIVNEVKGYYIKVDGKYVYVYLKDAHA 120

QY 145 DNVRTKDEINRQKEHVKNDEKVNNAVARSQGRYTTNDGYVFNPAITDTCNAVIVP 204

Db 121 DNVRTKDEINRQKEHVKNDEKVNNAVARSQGRYTTNDGYVFNPAITDTCNAVIVP 180

QY 205 HGGHYHVIKPSDLSASELAALAGKNMOPSOISYSSSTASDNTQSVAKGSTSKPANK 264

Db 181 HGGHYHVIKPSDLSASELAALAGKNMOPSOISYSSSTASDNTQSVAKGSTSKPANK 240

QY 265 SENLQSLKELYDPSAQRYSESDGLVDFDPAKIIISRTPNGVAIPHGDHYHFIPYSKLSAL 324

Db 241 SENLQSLKELYDPSAQRYSESDGLVDFDPAKIIISRTPNGVAIPHGDHYHFIPYSKLSAL 300

QY 325 EEKTAARVPISTGSTVSTNAKPNEVYSSLSGSLSSNPSSLTTSKELSSADGYYIFNPKDI 384

Db 301 EEKTAARVPISTGSTVSTNAKPNEVYSSLSGSLSSNPSSLTTSKELSSADGYYIFNPKDI 360

QY 385 VEETATAYIVRHGDHFYIPKSNQIGOPTLPNNSLATPSPSLNPCTSHKHEEDGYGF 444

Db 361 VEETATAYIVRHGDHFYIPKSNQIGOPTLPNNSLATPSPSLNPCTSHKHEEDGYGF 420

QY 445 DANRIIAEDSGFVMSHGDHNNHYEFKK 471

Db 421 DANRIIAEDSGFVMSHGDHNNHYEFKK 447

RESULT 2

US-08-961-083-56

; Sequence 56, Application US/08961083

; Patent No. 6159469

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-56

Query Match 3.1%; Score 32; DB 4; Length 796;
Best Local Similarity 100.0%; Pred. No. 7.1e-23;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 AEQIVKITDQGYVTSYHGHYHYNGKVPYDA 98
|||||
DB 46 AEQIVKITDQGYVTSYHGHYHYNGKVPYDA 77
|||||

RESULT 3
US-08-961-083-66
Sequence 66, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-66

Query Match 2.7%; Score 28; DB 4; Length 763;
Best Local Similarity 100.0%; Pred. No. 6.1e-19;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 VIKITDQGYVTSYHGHYHYNGKVPYDA 98
|||||
DB 52 VIKITDQGYVTSYHGHYHYNGKVPYDA 79
|||||

RESULT 4
US-09-049-577-2
Sequence 2, Application US/09049577
Patent No. 5968804
GENERAL INFORMATION:
APPLICANT: Feng, Teng-yung
APPLICANT: Lin, Hao-jan
TITLE OF INVENTION: AMPHIPATHIC PROTEIN-1
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/049,577
FILING DATE: 27-MAR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Tsao, Y. Rocky
REGISTRATION NUMBER: 34,053
REFERENCE/DOCKET NUMBER: 08919/007001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-049-577-2

Query Match 0.8%; Score 8; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
|||||
DB 52 VKLITPDG 59
|||||

RESULT 5

US-09-390-598-2
; Sequence 2, Application US/09390598
; Patent No. 6297360
; GENERAL INFORMATION:
; APPLICANT: Feng, Teng-yung
; APPLICANT: Lin, Hao-jan
; TITLE OF INVENTION: AMPHIPATHIC PROTEIN-1
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/390,598
; FILING DATE: 27-MAR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsao, Y. Rocky
; REGISTRATION NUMBER: 34,053
; REFERENCE/DOCKET NUMBER: 08919/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-390-598-2

Query Match 0.8%; Score 8; DB 4; Length 144;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 VKLITPDG 648
Db 52 VKLITPDG 59

Search completed: May 13, 2003, 14:01:59
Job time : 20 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 14:00:08 ; Search time 28 Seconds
(without alignments)
3414.811 Million cell updates/sec

Title: US-09-471-255-2

Perfect score: 1039

Sequence: 1 MKFSKYYIAGSAVIVLSL.....TELRLPSGEVINKNLSDFIA 1039

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 349150.seqs, 92025710 residues

Word size : 8

Total number of hits satisfying chosen parameters: 77

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 500 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1039	100.0	1039	9	US-09-884-465A-6
2	807	77.7	1152	9	US-09-884-465A-379
3	807	77.7	1238	9	US-09-884-465A-381
4	807	77.7	1365	9	US-09-884-465A-382
5	807	77.7	1378	9	US-09-884-465A-378
6	578	53.6	633	9	US-09-884-465A-350
7	575	53.3	633	9	US-09-884-465A-351
8	575	53.3	633	9	US-09-884-465A-361
9	573	53.1	627	9	US-09-884-465A-366
10	573	53.1	633	9	US-09-884-465A-367
11	573	53.1	894	9	US-09-884-465A-337
12	573	53.1	895	9	US-09-884-465A-347
13	573	53.1	900	9	US-09-884-465A-335
14	573	53.1	900	9	US-09-884-465A-339
15	573	53.1	900	9	US-09-884-465A-341
16	573	53.1	901	9	US-09-884-465A-343
17	573	53.1	901	9	US-09-884-465A-345
18	568	54.7	569	9	US-09-884-465A-235
19	568	54.7	569	9	US-09-884-465A-255

ALIGNMENTS

RESULT 1

US-09-884-465A-6
; Sequence 6, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie

20	568	54.7	913	9	US-09-884-465A-384	Sequence 384, App
21	568	54.7	999	9	US-09-884-465A-376	Sequence 376, App
22	568	54.7	999	9	US-09-884-465A-377	Sequence 377, App
23	568	54.7	1126	9	US-09-884-465A-383	Sequence 383, App
24	568	54.7	1139	9	US-09-884-465A-380	Sequence 380, App
25	531	51.1	633	9	US-09-884-465A-354	Sequence 354, App
26	489	47.1	633	9	US-09-884-465A-352	Sequence 352, App
27	489	47.1	633	9	US-09-884-465A-357	Sequence 357, App
28	489	47.1	633	9	US-09-884-465A-358	Sequence 358, App
29	489	47.1	633	9	US-09-884-465A-359	Sequence 359, App
30	489	47.1	633	9	US-09-884-465A-360	Sequence 360, App
31	489	47.1	900	9	US-09-884-465A-333	Sequence 333, App
32	489	47.1	900	9	US-09-884-465A-334	Sequence 334, App
33	489	47.1	906	9	US-09-884-465A-332	Sequence 332, App
34	489	47.1	906	9	US-09-884-465A-369	Sequence 369, App
35	489	47.1	906	9	US-09-884-465A-370	Sequence 370, App
36	489	47.1	906	9	US-09-884-465A-373	Sequence 373, App
37	486	46.8	569	9	US-09-884-465A-356	Sequence 356, App
38	486	46.8	569	9	US-09-884-465A-353	Sequence 353, App
39	486	46.8	569	9	US-09-884-465A-355	Sequence 355, App
40	486	46.8	569	9	US-09-884-465A-363	Sequence 363, App
41	486	46.8	569	9	US-09-884-465A-364	Sequence 364, App
42	486	46.8	569	9	US-09-884-465A-365	Sequence 365, App
43	486	46.8	569	9	US-09-884-465A-371	Sequence 371, App
44	486	46.8	569	9	US-09-884-465A-372	Sequence 372, App
45	484	46.6	569	9	US-09-884-465A-256	Sequence 256, App
46	484	46.6	621	9	US-09-884-465A-368	Sequence 368, App
47	484	46.6	627	9	US-09-884-465A-367	Sequence 367, App
48	484	46.6	633	9	US-09-884-465A-338	Sequence 338, App
49	484	46.6	633	9	US-09-884-465A-339	Sequence 339, App
50	484	46.6	888	9	US-09-884-465A-348	Sequence 348, App
51	484	46.6	889	9	US-09-884-465A-348	Sequence 348, App
52	484	46.6	894	9	US-09-884-465A-336	Sequence 336, App
53	484	46.6	894	9	US-09-884-465A-340	Sequence 340, App
54	484	46.6	894	9	US-09-884-465A-342	Sequence 342, App
55	484	46.6	895	9	US-09-884-465A-344	Sequence 344, App
56	475	45.7	840	9	US-09-884-465A-346	Sequence 346, App
57	447	43.0	447	10	US-09-765-272-182	Sequence 182, App
58	94	9.0	94	9	US-09-884-465A-11	Sequence 11, App
59	78	7.5	78	9	US-09-884-465A-19	Sequence 19, App
60	57	5.5	57	9	US-09-884-465A-20	Sequence 20, App
61	51	4.9	51	9	US-09-884-465A-12	Sequence 12, App
62	42	4.0	42	9	US-09-884-465A-18	Sequence 18, App
63	40	3.8	40	9	US-09-884-465A-13	Sequence 13, App
64	39	3.8	39	9	US-09-884-465A-14	Sequence 14, App
65	36	3.5	36	9	US-09-884-465A-16	Sequence 16, App
66	35	3.4	35	9	US-09-884-465A-16	Sequence 16, App
67	32	3.1	796	10	US-09-765-272-56	Sequence 56, App
68	32	3.1	838	9	US-09-884-465A-8	Sequence 8, App
69	32	3.1	840	9	US-09-884-465A-7	Sequence 7, App
70	28	2.7	763	10	US-09-765-272-66	Sequence 66, App
71	24	2.3	24	9	US-09-884-465A-15	Sequence 15, App
72	16	1.5	793	9	US-09-252-088-15	Sequence 15, App
73	8	0.8	543	9	US-09-746-783-78	Sequence 78, App
74	8	0.8	715	9	US-09-252-088-16	Sequence 16, App
75	8	0.8	1134	9	US-09-836-392-16	Sequence 16, App
76	8	0.8	1237	10	US-09-862-027-78	Sequence 78, App
77	8	0.8	1308	10	US-09-862-027-79	Sequence 79, App

APPLICANT: Ouellet, Catherine			
TITLE OF INVENTION: Streptococcus Antigens			
FILE REFERENCE: 055190-0044			
CURRENT APPLICATION NUMBER: US/09/884,465A			
CURRENT FILING DATE: 2001-06-20			
PRIOR APPLICATION NUMBER: 60/212,683			
PRIOR FILING DATE: 2000-06-20			
NUMBER OF SEQ ID NOS: 384			
SOFTWARE: PatentIn version 3.1			
SEQ ID NO 6			
LENGTH: 1039			
TYPE: PRT			
ORGANISM: Streptococcus pneumoniae			
US-09-884-465A-6			
Query Match		100.0%;	Score 1039; DB 9; Length 1039;
Best Local Similarity		100.0%;	Pred. No. 0;
Matches 1039; Conservative		0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	MKFSKYIAGSAVIVSLCAVALNQHRSQENKNNRVSYVDGSSOSSOKSENLTDPQVS	60
Db	1	MKFSKYIAGSAVIVSLCAVALNQHRSQENKNNRVSYVDGSSOSSOKSENLTDPQVS	60
Qy	61	QKSGIAEQIVIKITDQGYVTSRSHGHYHYNGKVPYDALFSEELLMKDPNYQIKDADIVN	120
Db	61	QKSGIAEQIVIKITDQGYVTSRSHGHYHYNGKVPYDALFSEELLMKDPNYQIKDADIVN	120
Qy	121	EVKGYIIVKGVYVYLKDAHADNVRKDEINROKQDHEVNDKRVNSNVAVARSQRY	180
Db	121	EVKGYIIVKGVYVYLKDAHADNVRKDEINROKQDHEVNDKRVNSNVAVARSQRY	180
Qy	181	TTNDGVYVNPADIIEDTGNAVIVPHGCHYHYIPKSDLSASELAHAHLAGNMPQSLS	240
Db	181	TTNDGVYVNPADIIEDTGNAVIVPHGCHYHYIPKSDLSASELAHAHLAGNMPQSLS	240
Qy	241	YSSTASDNNTQSVAKGSTSKPANKSENLSLKLKELYDPSAQRYSSESDGLVDPKAIISR	300
Db	241	YSSTASDNNTQSVAKGSTSKPANKSENLSLKLKELYDPSAQRYSSESDGLVDPKAIISR	300
Qy	301	TPNGVAIPHGDHYHFIPIYSKLSALEEKIARMPVISTGTVSTNAKPNEVWSLGSLSN	360
Db	301	TPNGVAIPHGDHYHFIPIYSKLSALEEKIARMPVISTGTVSTNAKPNEVWSLGSLSN	360
Qy	361	PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQICQPTLPNNSLA	420
Db	361	PSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQICQPTLPNNSLA	420
Qy	421	TPSPSLPDPNCTSHKHEEDGYGFDANRIIAEDSESGFVMSHGDHNYFPFKKOLTEQIKA	480
Db	421	TPSPSLPDPNCTSHKHEEDGYGFDANRIIAEDSESGFVMSHGDHNYFPFKKOLTEQIKA	480
Qy	481	AQKHEEVKTSHNGLDLSLSSHEODYPGNKEMKDLKKEEIKIAGIMKQYGVKRESIVVN	540
Db	481	AQKHEEVKTSHNGLDLSLSSHEODYPGNKEMKDLKKEEIKIAGIMKQYGVKRESIVVN	540
Qy	541	KEKNALIIYPHGDHHDADIDIDHAPVIGHSHSNLYFLFKPEEGVAKKEGKRYITGEELTNV	600
Db	541	KEKNALIIYPHGDHHDADIDIDHAPVIGHSHSNLYFLFKPEEGVAKKEGKRYITGEELTNV	600
Qy	601	VNLKSTFNQNTLANGOKRYSFSEFPPELEKLGINMLVKLITPDGKYLEKVSQKVF	660
Db	601	VNLKSTFNQNTLANGOKRYSFSEFPPELEKLGINMLVKLITPDGKYLEKVSQKVF	660
Qy	661	EGVGNIAINFELDQYPLGQTKYTIASKQYPEVSYDGTFTVPTPSLAYKMASQTIFFPFFIA	720
Db	661	EGVGNIAINFELDQYPLGQTKYTIASKQYPEVSYDGTFTVPTPSLAYKMASQTIFFPFFIA	720
Qy	721	GDTYLRVNPQFVAPKGDALVRVDEPHGNAYLENNYKVEIKLPIPKLNOGTTTRAGNK	780
Db	721	GDTYLRVNPQFVAPKGDALVRVDEPHGNAYLENNYKVEIKLPIPKLNOGTTTRAGNK	780
Qy	781	IPVTFMANAYLDNQSTYIVVEVPILEKENQTDKPSILPQFKRKAQENSKLDEKVEPKTS	840
Db	781	IPVTFMANAYLDNQSTYIVVEVPILEKENQTDKPSILPQFKRKAQENSKLDEKVEPKTS	840
Qy	841	EKVEKEKLSSETNSTLEEVPTVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLP	900
Db	841	EKVEKEKLSSETNSTLEEVPTVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLP	900
Qy	901	SCEVIKKNMADFTGEAPQNGENKPSKNGKSTGTCTVSNQPTENKPADSLPEAPNEKPKP	960
Db	901	SCEVIKKNMADFTGEAPQNGENKPSKNGKSTGTCTVSNQPTENKPADSLPEAPNEKPKP	960
Qy	961	ENSTDNGLNPEGNVSDPMLDPALEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTI	1020
Db	961	ENSTDNGLNPEGNVSDPMLDPALEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTI	1020
Qy	1021	ELRLPSGEVIKKNLSDFTA	1039
Db	1021	ELRLPSGEVIKKNLSDFTA	1039
RESULT 2			
US-09-884-465A-379			
; Sequence 379, Application US/09884465A			
; Publication No. US20030077293A1			
; GENERAL INFORMATION:			
; APPLICANT: Shire Biochem, Inc.			
; APPLICANT: Hamel, Josee			
; APPLICANT: Brodeur, Bernard			
; APPLICANT: Martin, Denis			
; APPLICANT: Charland, Nathalie			
; APPLICANT: Ouellet, Catherine			
; TITLE OF INVENTION: Streptococcus Antigens			
; FILE REFERENCE: 055190-0044			
; CURRENT APPLICATION NUMBER: US/09/884,465A			
; CURRENT FILING DATE: 2001-06-20			
; PRIOR APPLICATION NUMBER: 60/212,683			
; PRIOR FILING DATE: 2000-06-20			
; NUMBER OF SEQ ID NOS: 384			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 379			
; LENGTH: 1152			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Unknown Organism			
; NAME/KEY: MISC_FEATURE			
; LOCATION: (1)..(1)			
; OTHER INFORMATION: Xaa = Methionine or nothing			
; NAME/KEY: MISC_FEATURE			
; LOCATION: (344)..(344)			
; OTHER INFORMATION: Xaa = Glycine or nothing			
; NAME/KEY: MISC_FEATURE			
; LOCATION: (345)..(345)			
; OTHER INFORMATION: Xaa = Proline or nothing			
US-09-884-465A-379			
Query Match		77.7%;	Score 807; DB 9; Length 1152;
Best Local Similarity		100.0%;	Pred. No. 0;
Matches 807; Conservative		0;	Mismatches 0; Indels 0; Gaps 0;
Qy	233	NMQPSQLSYSTASDNNTQSVAKGSTSKPANKSENLSLKLKELYDPSAQRYSSESDGLVF	292
Db	346	NMQPSQLSYSTASDNNTQSVAKGSTSKPANKSENLSLKLKELYDPSAQRYSSESDGLVF	405
Qy	293	DPAKIISRTPNGVAIPHGDHYHFIPIYSKLSALEEKIARMPVISTGTVSTNAKPNEVWS	352
Db	406	DPAKIISRTPNGVAIPHGDHYHFIPIYSKLSALEEKIARMPVISTGTVSTNAKPNEVWS	465
Qy	353	SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQICQ	412
Db	466	SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQICQ	525
Qy	413	TLPNNSLATPSPSLPDPNCTSHKHEEDGYGFDANRIIAEDSESGFVMSHGDHNYFPFK	472

```
Db 526 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDESQFVMSHGDNHFFKKD 585
QY 473 LTEEQIKAAQKHLLEEVKTSHNGLDSSSHEQDYPGNKAKEMKDLDDKIEEKIAGIMKOYGV 532
Db 586 LTEEQIKAAQKHLLEEVKTSHNGLDSSSHEQDYPGNKAKEMKDLDDKIEEKIAGIMKOYGV 645
QY 533 KRESIVVNEKKNALIIYPHGDHHDADIDEHKPVGIGHSHSNVELPKPEGVAKKGNKYV 592
Db 646 KRESIVVNEKKNALIIYPHGDHHDADIDEHKPVGIGHSHSNVELPKPEGVAKKGNKYV 705
QY 593 TGEELNVVNLKNSFNQNTFLANGQKRVSFSPPELEKLGINMLVKLLTPDGKYLE 652
Db 706 TGEELNVVNLKNSFNQNTFLANGQKRVSFSPPELEKLGINMLVKLLTPDGKYLE 765
QY 653 KVSCKVFGCVGNIANFELDQYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASQ 712
Db 766 KVSCKVFGCVGNIANFELDQYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASQ 825
QY 713 TIFYPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGKIKLPIPKLNQ 772
Db 826 TIFYPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGKIKLPIPKLNQ 885
QY 773 TTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKENKAQENSKLDE 832
Db 886 TTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKENKAQENSKLDE 945
QY 833 KVEEPTSEKVEKEKLSSETNSTLSHEQDYPGNKAKEMKDLDDKIEEKIAGIMKOYGV 892
Db 946 KVEEPTSEKVEKEKLSSETNSTLSHEQDYPGNKAKEMKDLDDKIEEKIAGIMKOYGV 1005
QY 893 GTIELYLPSPGEVVIKKNMADFTGEAOGNGENKPSNGKYSTGTVENQPTENKPADSLPEA 952
Db 1006 GTIELYLPSPGEVVIKKNMADFTGEAOGNGENKPSNGKYSTGTVENQPTENKPADSLPEA 1065
QY 953 PNEKPVKPNSTDNGLNPEGVNGSDPMLDPALEAPAVDPVOEKLEKFTASYGLGLDSV 1012
Db 1066 PNEKPVKPNSTDNGLNPEGVNGSDPMLDPALEAPAVDPVOEKLEKFTASYGLGLDSV 1125
QY 1013 IFNMDGTIELRPSGEVVIKKNLSDFIA 1039
Db 1126 IFNMDGTIELRPSGEVVIKKNLSDFIA 1152
```

RESULT 3

```
US-09-884-465A-381
; Sequence 381, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 1238
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (430)..(430)
```

RESULT 4

```
US-09-884-465A-382
; Sequence 382, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
```

```
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (431)..(431)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-381
```

```
Query Match 77.7%; Score 807; DB 9; Length 1238;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 233 NMOPSQLSYSTASDNNTQSVAKGSTKSPANKSENQSLKELYDSPAQRYSDGLVLF 292
Db 432 NMOPSQLSYSTASDNNTQSVAKGSTKSPANKSENQSLKELYDSPAQRYSDGLVLF 491
QY 293 DPAKIIISITPNCVATPHGDHYHFFIPYSKLSALEEKIARMVPIISGTSTVSTNAKPNVVS 352
Db 492 DPAKIIISITPNCVATPHGDHYHFFIPYSKLSALEEKIARMVPIISGTSTVSTNAKPNVVS 551
QY 353 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIHYTPKSNQIQGP 412
Db 552 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIHYTPKSNQIQGP 611
QY 413 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDESQFVMSHGDNHFFKKD 472
Db 612 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDESQFVMSHGDNHFFKKD 671
QY 473 LTEEQIKAAQKHLLEEVKTSHNGLDSSSHEQDYPGNKAKEMKDLDDKIEEKIAGIMKOYGV 532
Db 672 LTEEQIKAAQKHLLEEVKTSHNGLDSSSHEQDYPGNKAKEMKDLDDKIEEKIAGIMKOYGV 731
QY 533 KRESIVVNEKKNALIIYPHGDHHDADIDEHKPVGIGHSHSNVELPKPEGVAKKGNKYV 592
Db 732 KRESIVVNEKKNALIIYPHGDHHDADIDEHKPVGIGHSHSNVELPKPEGVAKKGNKYV 791
QY 593 TGEELNVVNLKNSFNQNTFLANGQKRVSFSPPELEKLGINMLVKLLTPDGKYLE 652
Db 792 TGEELNVVNLKNSFNQNTFLANGQKRVSFSPPELEKLGINMLVKLLTPDGKYLE 851
QY 653 KVSCKVFGCVGNIANFELDQYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASQ 712
Db 852 KVSCKVFGCVGNIANFELDQYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASQ 911
QY 713 TIFYPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGKIKLPIPKLNQ 772
Db 912 TIFYPHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGKIKLPIPKLNQ 971
QY 773 TTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKENKAQENSKLDE 832
Db 972 TTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKENKAQENSKLDE 1031
QY 833 KVEEPTSEKVEKEKLSSETNSTLSHEQDYPGNKAKEMKDLDDKIEEKIAGIMKOYGV 892
Db 1032 KVEEPTSEKVEKEKLSSETNSTLSHEQDYPGNKAKEMKDLDDKIEEKIAGIMKOYGV 1091
QY 893 GTIELYLPSPGEVVIKKNMADFTGEAOGNGENKPSNGKYSTGTVENQPTENKPADSLPEA 952
Db 1092 GTIELYLPSPGEVVIKKNMADFTGEAOGNGENKPSNGKYSTGTVENQPTENKPADSLPEA 1151
QY 953 PNEKPVKPNSTDNGLNPEGVNGSDPMLDPALEAPAVDPVOEKLEKFTASYGLGLDSV 1012
Db 1152 PNEKPVKPNSTDNGLNPEGVNGSDPMLDPALEAPAVDPVOEKLEKFTASYGLGLDSV 1211
QY 1013 IFNMDGTIELRPSGEVVIKKNLSDFIA 1039
Db 1212 IFNMDGTIELRPSGEVVIKKNLSDFIA 1238
```

APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 382
LENGTH: 1365
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
NAME/KEY: MISC_FEATURE
LOCATION: (1)..(1)
OTHER INFORMATION: Xaa = Methionine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (557)..(557)
OTHER INFORMATION: Xaa = Glycine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (558)..(558)
OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-382

Query Match 77.7%; Score 807; DB 9; Length 1365;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NMQPSQLSYSTASDNNNTQSVAKGSTSKPANKSENQLSKELYDPSAQRYSSESGLVF 292
DB 559 NMQPSQLSYSTASDNNNTQSVAKGSTSKPANKSENQLSKELYDPSAQRYSSESGLVF 618
QY 293 DPAKIIERTPNGVAIPGHDHYHPIFYPSKLSALEEKIARMPVPSGTSTVSTNAKNEVVS 352
DB 619 DPAKIIERTPNGVAIPGHDHYHPIFYPSKLSALEEKIARMPVPSGTSTVSTNAKNEVVS 578
QY 353 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOP 412
DB 679 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOP 738
QY 413 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNFHFFKDD 472
DB 739 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNFHFFKDD 798
QY 473 LTERQIRAAQKHLSEVTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOP 532
DB 799 LTERQIRAAQKHLSEVTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOP 858
QY 533 KRESIVNKEKNALIIYPHGDHHDADPIDEHKKPVGICHSHSNVELPPEGVAKKEGKYY 592
DB 859 KRESIVNKEKNALIIYPHGDHHDADPIDEHKKPVGICHSHSNVELPPEGVAKKEGKYY 918
QY 593 TGEELTNVNLKSTNNFTLANGQKRVSFSPFPELEKKGILNMLVKLITPDGKYLE 652
DB 919 TGEELTNVNLKSTNNFTLANGQKRVSFSPFPELEKKGILNMLVKLITPDGKYLE 978
QY 653 KVSQKVGEGVGNANTANFELDQPLPGQTFYTTASXDYFEVSDGTFYPTSLAYKMASQ 712
DB 979 KVSQKVGEGVGNANTANFELDQPLPGQTFYTTASXDYFEVSDGTFYPTSLAYKMASQ 1038
QY 713 TIFYPFHAGDTYLRVNPQFAPVPGTDAIVRVDFEFGHNAVLENNYKVGKIKPIPKLNOG 772
DB 1039 TIFYPFHAGDTYLRVNPQFAPVPGTDAIVRVDFEFGHNAVLENNYKVGKIKPIPKLNOG 1098
QY 773 TTRTAGNKIPVTFMANAYLONQSTYIVVEPILEKENOTKPSILPOFKNKAOENSKLDE 832
DB 1099 TTRTAGNKIPVTFMANAYLONQSTYIVVEPILEKENOTKPSILPOFKNKAOENSKLDE 1158

QY 833 KVEEPKTSKVEKEKLETSNSTNSTLEEVPTVDVQEKVAKFAESYGMKLENVLFNMD 892
DB 1159 KVEEPKTSKVEKEKLETSNSTNSTLEEVPTVDVQEKVAKFAESYGMKLENVLFNMD 1218
QY 893 GTIELYLPSEGEVYKKNWADTGEAPQNGENKPSSENGKYSTGVENQPTENKPADSLPEA 952
DB 1219 GTIELYLPSEGEVYKKNWADTGEAPQNGENKPSSENGKYSTGVENQPTENKPADSLPEA 1278
QY 953 PNEKPVKPEKSTNGMLNPEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1012
DB 1279 PNEKPVKPEKSTNGMLNPEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1338
QY 1013 IFNMDGTIELRLPSGEVYKKNLSDFIA 1039
DB 1339 IFNMDGTIELRLPSGEVYKKNLSDFIA 1365

RESULT 5

US-09-884-465A-378
Sequence 378, Application US/09884465A
Publication No US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 378
LENGTH: 1378
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
NAME/KEY: MISC_FEATURE
LOCATION: (1)..(1)
OTHER INFORMATION: Xaa = Methionine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (570)..(570)
OTHER INFORMATION: Xaa = Glycine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (571)..(571)
OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-378

Query Match 77.7%; Score 807; DB 9; Length 1378;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 NMQPSQLSYSTASDNNNTQSVAKGSTSKPANKSENQLSKELYDPSAQRYSSESGLVF 292
DB 572 NMQPSQLSYSTASDNNNTQSVAKGSTSKPANKSENQLSKELYDPSAQRYSSESGLVF 631
QY 293 DPAKIIERTPNGVAIPGHDHYHPIFYPSKLSALEEKIARMPVPSGTSTVSTNAKNEVVS 352
DB 632 DPAKIIERTPNGVAIPGHDHYHPIFYPSKLSALEEKIARMPVPSGTSTVSTNAKNEVVS 691
QY 353 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOP 412
DB 692 SLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKSNQIGOP 751
QY 413 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNFHFFKDD 472
DB 752 TLPNNSLATPSPSLPDPGTSHEKHEEDGYGFDANRIIAEDSGFVMSHGDHNFHFFKDD 811

QY 473 L TEEQIKAAQHLEEVKTSNGLDLSLSSHEQDYPGNKEMKMDLKKIEKIAGIMQYGV 532
D 812 L TEEQIKAAQHLEEVKTSNGLDLSLSSHEQDYPGNKEMKMDLKKIEKIAGIMQYGV 871
QY 533 KRESIVVAKENNAIIPYPGDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKGNKVY 592
D 872 KRESIVVAKENNAIIPYPGDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKGNKVY 931
QY 593 TGEELTNVNLKSTFNQNTLANGOKRVSFSPFPPELEKKGINMLVKLITPDGKYLE 652
D 932 TGEELTNVNLKSTFNQNTLANGOKRVSFSPFPPELEKKGINMLVKLITPDGKYLE 991
QY 653 KVSQVFGEGVGNIANFELDPQLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQ 712
D 992 KVSQVFGEGVGNIANFELDPQLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQ 1051
QY 713 TIFYPFPHAGDTYLRVNPQFAVPKGTDALRVFDEFHGNAYLENNYKVGGEIKLPKLNQ 772
D 1052 TIFYPFPHAGDTYLRVNPQFAVPKGTDALRVFDEFHGNAYLENNYKVGGEIKLPKLNQ 1111
QY 773 TTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKQENQTDKPSILPQFKRNKAQENSKLDE 832
D 1112 TTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKQENQTDKPSILPQFKRNKAQENSKLDE 1171
QY 833 KVEEPTSEKVEKEKLSGTGNSTNSSTLEEYPTDVPQEKVAKPAESYGMKLENVLFNMD 892
D 1172 KVEEPTSEKVEKEKLSGTGNSTNSSTLEEYPTDVPQEKVAKPAESYGMKLENVLFNMD 1231
QY 893 GTIELYPSGSEVIKKNMADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPEA 952
D 1232 GTIELYPSGSEVIKKNMADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPEA 1291
QY 953 PNEKPVKPNSTNDGMLNPEGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1012
D 1292 PNEKPVKPNSTNDGMLNPEGVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSV 1351
QY 1013 IFNMDGTIELRPSGEVIKKNLSDFIA 1039
D 1352 IFNMDGTIELRPSGEVIKKNLSDFIA 1378

RESULT 6

US-09-884-465A-350
; Sequence 350, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 350
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-350

Query Match 55.6%; Score 578; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 GDHNHYFFKDLTEEQIKAAQHLEEVKTSNGLDLSLSSHEQDYPGNKEMKMDLKKIEE 521
D 56 GDHNHYFFKDLTEEQIKAAQHLEEVKTSNGLDLSLSSHEQDYPGNKEMKMDLKKIEE 115
QY 522 KIAGIMQYGVKRESIVVAKENNAIIPYPGDHHADPIDEHKPVGIGHSHSNYELFKPEE 561
D 116 KIAGIMQYGVKRESIVVAKENNAIIPYPGDHHADPIDEHKPVGIGHSHSNYELFKPEE 175
QY 582 GVAKKEGNKYVTGEBELTNVNLKSTFNQNTLANGOKRVSFSPFPPELEKKGINMLV 641
D 176 GVAKKEGNKYVTGEBELTNVNLKSTFNQNTLANGOKRVSFSPFPPELEKKGINMLV 235
QY 642 KLITPDGKYLEKVSQVFGEGVGNIANFELDPQLPGQTFKYTIASKDYPEVSYDGTFTV 701
D 236 KLITPDGKYLEKVSQVFGEGVGNIANFELDPQLPGQTFKYTIASKDYPEVSYDGTFTV 295
QY 702 PTSLAYKMASQTFYPPHAGDTYLRVNPQFAVPKGTDALRVFDEFHGNAYLENNYKVG 761
D 296 PTSLAYKMASQTFYPPHAGDTYLRVNPQFAVPKGTDALRVFDEFHGNAYLENNYKVG 355
QY 762 IKPIPLKNGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKQENQTDKPSILPQFKR 821
D 356 IKPIPLKNGTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKQENQTDKPSILPQFKR 415
QY 822 NKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSTNSSTLEEYPTDVPQEKVAKPAESY 881
D 416 NKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSTNSSTLEEYPTDVPQEKVAKPAESY 475
QY 882 MKLENVLFNMDGTIELYPSGSEVIKKNMADFTGEAPQNGENKPSNGKSVSTGTVENQPT 941
D 476 MKLENVLFNMDGTIELYPSGSEVIKKNMADFTGEAPQNGENKPSNGKSVSTGTVENQPT 535
QY 942 ENKPADSLPEAPNEKPVKPNSTNDGMLNPEGVSDPMLDPALEAPAVDPVQEKLEK 1001
D 536 ENKPADSLPEAPNEKPVKPNSTNDGMLNPEGVSDPMLDPALEAPAVDPVQEKLEK 595
QY 1002 TASYGLGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
D 596 TASYGLGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 633

RESULT 7

US-09-884-465A-351
; Sequence 351, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 351
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-351

Query Match 55.3%; Score 575; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 NHYFFKDLTEEQIKAAQHLEEVKTSNGLDLSLSSHEQDYPGNKEMKMDLKKIEKIA 524

Db 59 NHYFFKDLTEEQIAAKQHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDDKKIEEKIA 118
QY 525 GIMKOYGVKRESIVNKEKNAIYPHGDHHDADPIDSHKPVGIGHSHSNYELFKPEEGVA 584
Db 119 GIMKOYGVKRESIVNKEKNAIYPHGDHHDADPIDSHKPVGIGHSHSNYELFKPEEGVA 178
QY 585 KKEGNKYVTGEELTNVNLKSNFTNQNFTLANGQKRVSFSPPELEKKGILNMLVKLI 644
Db 179 KKEGNKYVTGEELTNVNLKSNFTNQNFTLANGQKRVSFSPPELEKKGILNMLVKLI 238
QY 645 TPDGKLVLEKSVGKVFEGVGNIANFELDPYLPQOTFKYTIASKDYPEVSYDGTFTVPTS 704
Db 239 TPDGKLVLEKSVGKVFEGVGNIANFELDPYLPQOTFKYTIASKDYPEVSYDGTFTVPTS 298
QY 705 LAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNKYVGEIKL 764
Db 299 LAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNKYVGEIKL 358
QY 765 PIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYVEVPILEKENQTDKPSILPQKRNKA 824
Db 359 PIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYVEVPILEKENQTDKPSILPQKRNKA 418
QY 825 QENSKLDEKVEEPTSEKVEKEKLSGTNSTLEEVPTVDPVQEKVAKFAESYGMKL 884
Db 419 QENSKLDEKVEEPTSEKVEKEKLSGTNSTLEEVPTVDPVQEKVAKFAESYGMKL 478
QY 885 ENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKSVGTGTVENQPTENK 944
Db 479 ENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKSVGTGTVENQPTENK 538
QY 945 PADSLPEAPNEKPKVPENSTNDGMLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTAS 1004
Db 539 PADSLPEAPNEKPKVPENSTNDGMLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTAS 598
QY 1005 YGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 599 YGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 633

RESULT 8

US-09-884-465A-361
; Sequence 361, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 361
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-361

Query Match 55.3%; Score 575; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 NHYFFKDLTEEQIAAKQHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDDKKIEEKIA 524
Db 55 YFFKDLTEEQIAAKQHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDDKKIEEKIAGI 114

Db 59 NHYFFKDLTEEQIAAKQHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDDKKIEEKIA 118
QY 525 GIMKOYGVKRESIVNKEKNAIYPHGDHHDADPIDSHKPVGIGHSHSNYELFKPEEGVA 584
Db 119 GIMKOYGVKRESIVNKEKNAIYPHGDHHDADPIDSHKPVGIGHSHSNYELFKPEEGVA 178
QY 585 KKEGNKYVTGEELTNVNLKSNFTNQNFTLANGQKRVSFSPPELEKKGILNMLVKLI 644
Db 179 KKEGNKYVTGEELTNVNLKSNFTNQNFTLANGQKRVSFSPPELEKKGILNMLVKLI 238
QY 645 TPDGKLVLEKSVGKVFEGVGNIANFELDPYLPQOTFKYTIASKDYPEVSYDGTFTVPTS 704
Db 239 TPDGKLVLEKSVGKVFEGVGNIANFELDPYLPQOTFKYTIASKDYPEVSYDGTFTVPTS 298
QY 705 LAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNKYVGEIKL 764
Db 299 LAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNKYVGEIKL 358
QY 765 PIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYVEVPILEKENQTDKPSILPQKRNKA 824
Db 359 PIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYVEVPILEKENQTDKPSILPQKRNKA 418
QY 825 QENSKLDEKVEEPTSEKVEKEKLSGTNSTLEEVPTVDPVQEKVAKFAESYGMKL 884
Db 419 QENSKLDEKVEEPTSEKVEKEKLSGTNSTLEEVPTVDPVQEKVAKFAESYGMKL 478
QY 885 ENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKSVGTGTVENQPTENK 944
Db 479 ENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKSVGTGTVENQPTENK 538
QY 945 PADSLPEAPNEKPKVPENSTNDGMLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTAS 1004
Db 539 PADSLPEAPNEKPKVPENSTNDGMLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTAS 598
QY 1005 YGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 599 YGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 633

RESULT 9

US-09-884-465A-366
; Sequence 366, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 366
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-366

Query Match 55.1%; Score 573; DB 9; Length 627;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKDLTEEQIAAKQHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDDKKIEEKIAGI 526
Db 55 YFFKDLTEEQIAAKQHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDDKKIEEKIAGI 114

QY 527 MKQYGVKRESIVVNEKNNAIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586
DB 115 MKQYGVKRESIVVNEKNNAIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 174
QY 587 EGNKYVTGEEELTNVNNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINMLVKLITP 646
DB 175 EGNKYVTGEEELTNVNNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINMLVKLITP 234
QY 647 DGKVLKSVKGVFEGVGNIANFELDQYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLA 706
DB 235 DGKVLKSVKGVFEGVGNIANFELDQYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLA 294
QY 707 YKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGEEKLPI 766
DB 295 YKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGEEKLPI 354
QY 767 PKLNOGTRTAGNKIPVTFMANAYLDNOSTYIIVEPILKEKENQTDKPSILPQKRNKQAE 826
DB 355 PKLNOGTRTAGNKIPVTFMANAYLDNOSTYIIVEPILKEKENQTDKPSILPQKRNKQAE 414
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 886
DB 415 NSKLDEKVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 474
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKVTGTVENQPTENKPA 946
DB 475 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKVTGTVENQPTENKPA 534
QY 947 DSLPEAPNEKPKVPKPNSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
DB 535 DSLPEAPNEKPKVPKPNSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 594
QY 1007 LGLDSVFNMDGTIELRPSGEVIKKNLSDFIA 1039
DB 595 LGLDSVFNMDGTIELRPSGEVIKKNLSDFIA 627

RESULT 10

US-09-884-465A-362
; Sequence 362, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 362
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-362

Query Match 55.1%; Score 573; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKKDLTEEQIAAKQKHLVEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDRKKEIKIAGI 526
DB 61 YFFKKDLTEEQIAAKQKHLVEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDRKKEIKIAGI 120

QY 527 MKQYGVKRESIVVNEKNNAIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586
DB 121 MKQYGVKRESIVVNEKNNAIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 180
QY 587 EGNKYVTGEEELTNVNNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINMLVKLITP 646
DB 181 EGNKYVTGEEELTNVNNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINMLVKLITP 240
QY 647 DGKVLKSVKGVFEGVGNIANFELDQYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLA 706
DB 241 DGKVLKSVKGVFEGVGNIANFELDQYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLA 300
QY 707 YKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGEEKLPI 766
DB 301 YKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGHGNAYLENNYKVGEEKLPI 360
QY 767 PKLNOGTRTAGNKIPVTFMANAYLDNOSTYIIVEPILKEKENQTDKPSILPQKRNKQAE 826
DB 361 PKLNOGTRTAGNKIPVTFMANAYLDNOSTYIIVEPILKEKENQTDKPSILPQKRNKQAE 420
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 886
DB 421 NSKLDEKVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 480
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKVTGTVENQPTENKPA 946
DB 481 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSNGKVTGTVENQPTENKPA 540
QY 947 DSLPEAPNEKPKVPKPNSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
DB 541 DSLPEAPNEKPKVPKPNSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 600
QY 1007 LGLDSVFNMDGTIELRPSGEVIKKNLSDFIA 1039
DB 601 LGLDSVFNMDGTIELRPSGEVIKKNLSDFIA 633

RESULT 11

US-09-884-465A-337
; Sequence 337, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 337
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-337

Query Match 55.1%; Score 573; DB 9; Length 894;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKKDLTEEQIAAKQKHLVEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDRKKEIKIAGI 526
DB 322 YFFKKDLTEEQIAAKQKHLVEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLDRKKEIKIAGI 381
QY 527 MKQYGVKRESIVVNEKNNAIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586

Db 382 MKQGVKRESIVVWKEKNALIIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 441
QY 587 EGNKYVTGEBELTNVNVLLKNSFTNNQNTFLANGQKRVSFSPPELEKKLGINMLVKLITP 646
Db 442 EGNKYVTGEBELTNVNVLLKNSFTNNQNTFLANGQKRVSFSPPELEKKLGINMLVKLITP 501
QY 647 DGKLVKSVKGVGEGVGNIANFELDQPYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLA 706
Db 502 DGKLVKSVKGVGEGVGNIANFELDQPYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLA 561
QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGEIKLPI 766
Db 562 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGEIKLPI 621
QY 767 PKLNOGTTTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRKAQE 826
Db 622 PKLNOGTTTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRKAQE 681
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTNSTLSLSEVPTVDPVQEKVAKFAESYGMKLEN 886
Db 682 NSKLDEKVEEPTSEKVEKEKLSGTNSTLSLSEVPTVDPVQEKVAKFAESYGMKLEN 741
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKYSTGTVENOPTENKPA 946
Db 742 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKYSTGTVENOPTENKPA 801
QY 947 DSLPEAPNEKPKVPENSTONGMLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 802 DSLPEAPNEKPKVPENSTONGMLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 861
QY 1007 LGLDVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
Db 862 LGLDVIFNMDGTIELRPSGEVIKKNLSDFIA 894

RESULT 12
US-09-884-465A-347
; Sequence 347, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 347
; LENGTH: 895
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-347

Query Match 55.1%; Score 573; DB 9; Length 895;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKDLTEQIKAAQKHLSEVKTSHNGLDLSLSSHEQDYPGNAKEMKDLKKIEEKIAGI 526
Db 323 YFFKDLTEQIKAAQKHLSEVKTSHNGLDLSLSSHEQDYPGNAKEMKDLKKIEEKIAGI 382
QY 527 MKQGVKRESIVVWKEKNALIIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586

Db 383 MKQGVKRESIVVWKEKNALIIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 442
QY 587 EGNKYVTGEBELTNVNVLLKNSFTNNQNTFLANGQKRVSFSPPELEKKLGINMLVKLITP 646
Db 443 EGNKYVTGEBELTNVNVLLKNSFTNNQNTFLANGQKRVSFSPPELEKKLGINMLVKLITP 502
QY 647 DGKLVKSVKGVGEGVGNIANFELDQPYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLA 706
Db 503 DGKLVKSVKGVGEGVGNIANFELDQPYLPQGTFFKYTIASKDYPEVSYDGTFTVPTSLA 562
QY 707 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGEIKLPI 766
Db 563 YKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGEIKLPI 622
QY 767 PKLNOGTTTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRKAQE 826
Db 623 PKLNOGTTTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTDKPSILPQFKRKAQE 682
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTNSTLSLSEVPTVDPVQEKVAKFAESYGMKLEN 886
Db 683 NSKLDEKVEEPTSEKVEKEKLSGTNSTLSLSEVPTVDPVQEKVAKFAESYGMKLEN 742
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKYSTGTVENOPTENKPA 946
Db 743 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPOGNGENKPSNGKYSTGTVENOPTENKPA 802
QY 947 DSLPEAPNEKPKVPENSTONGMLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 803 DSLPEAPNEKPKVPENSTONGMLNPEGNGVSDPMLDPALEAPAVDPVQEKLEKFTASYG 862
QY 1007 LGLDVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
Db 863 LGLDVIFNMDGTIELRPSGEVIKKNLSDFIA 895

RESULT 13
US-09-884-465A-335
; Sequence 335, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 335
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-335

Query Match 55.1%; Score 573; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 467 YFFKDLTEQIKAAQKHLSEVKTSHNGLDLSLSSHEQDYPGNAKEMKDLKKIEEKIAGI 526
Db 328 YFFKDLTEQIKAAQKHLSEVKTSHNGLDLSLSSHEQDYPGNAKEMKDLKKIEEKIAGI 387
QY 527 MKQGVKRESIVVWKEKNALIIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586
Db 388 MKQGVKRESIVVWKEKNALIIYPHGDDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 447

```
QY 587 EGNKYVTGEELTNVNLKKNSTFNNQNTFLANGQKRVSEFPPPELEKKGILNMLVKLITP 646
DB 448 EGNKYVTGEELTNVNLKKNSTFNNQNTFLANGQKRVSEFPPPELEKKGILNMLVKLITP 507
QY 647 DGKLVKSVGKVFGEVGNIANFELDPQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLA 706
DB 508 DGKLVKSVGKVFGEVGNIANFELDPQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLA 567
QY 707 YKMASQTIIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGEIKLPI 766
DB 568 YKMASQTIIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGEIKLPI 627
QY 767 PKLNOGTTTAGNKPVTFMANAYLDNQSTIYVEVPILEKENQTDKPSILPQKRNKAQE 826
DB 628 PKLNOGTTTAGNKPVTFMANAYLDNQSTIYVEVPILEKENQTDKPSILPQKRNKAQE 687
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 886
DB 688 NSKLDEKVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 747
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSSENGKYSTGTVENQPTENKPA 946
DB 748 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSSENGKYSTGTVENQPTENKPA 807
QY 947 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
DB 808 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 867
QY 1007 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
DB 868 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 900
```

RESULT 14

```
US-09-884-465A-339
; Sequence 339, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 339
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-339
```

```
Query Match 55.1%; Score 573; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 467 YFFKKDLTEEQIKAAQKHLEEVKTSNGLDSLSSHEQDYPGNAXEMKDLDRKIEKTAGI 526
DB 328 YFFKKDLTEEQIKAAQKHLEEVKTSNGLDSLSSHEQDYPGNAXEMKDLDRKIEKTAGI 387
QY 527 MKQYGVKRESIWNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586
DB 388 MKQYGVKRESIWNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 447
```

```
QY 587 EGNKYVTGEELTNVNLKKNSTFNNQNTFLANGQKRVSEFPPPELEKKGILNMLVKLITP 646
DB 448 EGNKYVTGEELTNVNLKKNSTFNNQNTFLANGQKRVSEFPPPELEKKGILNMLVKLITP 507
QY 647 DGKLVKSVGKVFGEVGNIANFELDPQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLA 706
DB 508 DGKLVKSVGKVFGEVGNIANFELDPQYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLA 567
QY 707 YKMASQTIIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGEIKLPI 766
DB 568 YKMASQTIIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDEPHGNAYLENNYKVGEIKLPI 627
QY 767 PKLNOGTTTAGNKPVTFMANAYLDNQSTIYVEVPILEKENQTDKPSILPQKRNKAQE 826
DB 628 PKLNOGTTTAGNKPVTFMANAYLDNQSTIYVEVPILEKENQTDKPSILPQKRNKAQE 687
QY 827 NSKLDEKVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 886
DB 688 NSKLDEKVEEPTSEKVEKEKLSGTNSTSTLEEVPTVDPVQEKVAKFAESYGMKLEN 747
QY 887 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSSENGKYSTGTVENQPTENKPA 946
DB 748 VLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSSENGKYSTGTVENQPTENKPA 807
QY 947 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
DB 808 DSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 867
QY 1007 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039
DB 868 LGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 900
```

RESULT 15

```
US-09-884-465A-341
; Sequence 341, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 341
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-341
```

```
Query Match 55.1%; Score 573; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 467 YFFKKDLTEEQIKAAQKHLEEVKTSNGLDSLSSHEQDYPGNAXEMKDLDRKIEKTAGI 526
DB 328 YFFKKDLTEEQIKAAQKHLEEVKTSNGLDSLSSHEQDYPGNAXEMKDLDRKIEKTAGI 387
QY 527 MKQYGVKRESIWNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586
DB 388 MKQYGVKRESIWNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 447
QY 587 EGNKYVTGEELTNVNLKKNSTFNNQNTFLANGQKRVSEFPPPELEKKGILNMLVKLITP 646
```

Db 448 EGNKYVTGEELTNVWLLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 507
QY 647 DGKVLKSVKGVFEGVGNIANFELDPQYLPQGTFFKTYTIAASKDYPEVSYDGTFTVPTSLA 706
Db 509 DGKVLKSVKGVFEGVGNIANFELDPQYLPQGTFFKTYTIAASKDYPEVSYDGTFTVPTSLA 568
QY 707 YKASOTIFYPPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 766
Db 569 YKASOTIFYPPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 628
QY 767 PKLNGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQTDKPSILPQFKRNKAQE 826
Db 629 PKLNGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQTDKPSILPQFKRNKAQE 688
QY 827 NSKLDEKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTVDPVQEKVAKFAESYGMKLEN 886
Db 689 NSKLDEKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTVDPVQEKVAKFAESYGMKLEN 748
QY 887 VLFNMDGTIELYLPSEGEVKKNNADFTGEAPOGNGENKPSSENGKSVSTGVENQPTENKPA 946
Db 749 VLFNMDGTIELYLPSEGEVKKNNADFTGEAPOGNGENKPSSENGKSVSTGVENQPTENKPA 808
QY 947 DSLPEAPNEKPKPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 809 DSLPEAPNEKPKPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 868
QY 1007 LGDSVIFNMDGTIELRLPSGEVKKNNLSDFIA 1039
Db 869 LGDSVIFNMDGTIELRLPSGEVKKNNLSDFIA 901

RESULT 17
US-09-884-465A-345
; Sequence 345, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 345
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-345

Query Match 55.1%; Score 573; DB 9; Length 901;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 467 YFFKKDLTEQIKAAQKHLEEVKTSNGLDSSLSSHPQDYPGNAKEMKOLDKRIEIKIAGI 526
Db 329 YFFKKDLTEQIKAAQKHLEEVKTSNGLDSSLSSHPQDYPGNAKEMKOLDKRIEIKIAGI 388
QY 527 MKQYGVKRESIVVNNKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586
Db 389 MKQYGVKRESIVVNNKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 448
QY 587 EGNKYVTGEELTNVWLLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 646
Db 449 EGNKYVTGEELTNVWLLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 508

Db 449 EGNKYVTGEELTNVWLLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 508
QY 647 DGKVLKSVKGVFEGVGNIANFELDPQYLPQGTFFKTYTIAASKDYPEVSYDGTFTVPTSLA 706
Db 509 DGKVLKSVKGVFEGVGNIANFELDPQYLPQGTFFKTYTIAASKDYPEVSYDGTFTVPTSLA 568
QY 707 YKASOTIFYPPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 766
Db 569 YKASOTIFYPPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 628
QY 767 PKLNGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQTDKPSILPQFKRNKAQE 826
Db 629 PKLNGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQTDKPSILPQFKRNKAQE 688
QY 827 NSKLDEKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTVDPVQEKVAKFAESYGMKLEN 886
Db 689 NSKLDEKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTVDPVQEKVAKFAESYGMKLEN 748
QY 887 VLFNMDGTIELYLPSEGEVKKNNADFTGEAPOGNGENKPSSENGKSVSTGVENQPTENKPA 946
Db 749 VLFNMDGTIELYLPSEGEVKKNNADFTGEAPOGNGENKPSSENGKSVSTGVENQPTENKPA 808
QY 947 DSLPEAPNEKPKPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 1006
Db 809 DSLPEAPNEKPKPENSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYG 868
QY 1007 LGDSVIFNMDGTIELRLPSGEVKKNNLSDFIA 1039
Db 869 LGDSVIFNMDGTIELRLPSGEVKKNNLSDFIA 901

RESULT 17
US-09-884-465A-345
; Sequence 345, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 345
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-345

Query Match 55.1%; Score 573; DB 9; Length 901;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 467 YFFKKDLTEQIKAAQKHLEEVKTSNGLDSSLSSHPQDYPGNAKEMKOLDKRIEIKIAGI 526
Db 329 YFFKKDLTEQIKAAQKHLEEVKTSNGLDSSLSSHPQDYPGNAKEMKOLDKRIEIKIAGI 388
QY 527 MKQYGVKRESIVVNNKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 586
Db 389 MKQYGVKRESIVVNNKKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKK 448
QY 587 EGNKYVTGEELTNVWLLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 646
Db 449 EGNKYVTGEELTNVWLLKNSTFNQNFLLANGOKRVSFSPPELEKLGINMLVKLTIP 508

QY 647 DKVLEKVGKVGEGVGNIANFELDPQYLPQGTQKTYTIASKDYPEVSDGTFTVPTSLA 706
|||||
Db 509 DGKVLKVGKVGEGVGNIANFELDPQYLPQGTQKTYTIASKDYPEVSDGTFTVPTSLA 568
QY 707 YKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNAYLENNYKVGGEIKLPI 766
|||||
Db 569 YKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNAYLENNYKVGGEIKLPI 628
QY 767 PKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEVPILEKENQTDKPSILPOFKRNKAQE 826
|||||
Db 629 PKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIIVEVPILEKENQTDKPSILPOFKRNKAQE 688
QY 827 NSKLEKVEEPTSKVEKEKLSNLTGNSNLTLEEVPTDVPQVKAFABSYGKMLN 886
|||||
Db 689 NSKLEKVEEPTSKVEKEKLSNLTGNSNLTLEEVPTDVPQVKAFABSYGKMLN 748
QY 887 VLFNMDGTIELYLPSEGEVKKNNMADFTGEAPOGNGENKPSSENGKYSTGTVENQPTENKPA 946
|||||
Db 749 VLFNMDGTIELYLPSEGEVKKNNMADFTGEAPOGNGENKPSSENGKYSTGTVENQPTENKPA 808
QY 947 DSLPEAPNEKPKPNSNDGMLNPEGNVGSDPMLDPALEAPAVDPVQEKLEKTASYG 1006
|||||
Db - 809 DSLPEAPNEKPKPNSNDGMLNPEGNVGSDPMLDPALEAPAVDPVQEKLEKTASYG 868
QY 1007 LGLSDSVIFNMGTIELRLPSGEVKKNNLSDFIA 1039
|||||
Db 869 LGLSDSVIFNMGTIELRLPSGEVKKNNLSDFIA 901

RESULT 18

US-09-884-465A-235
; Sequence 235, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 235
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-235

Query Match 54.7%; Score 568; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLDDKKEEKTAGIMKQYG 531
|||||
Db 2 DLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLDDKKEEKTAGIMKQYG 61
QY 532 VKRESIVVNKEKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 591
|||||
Db 62 VKRESIVVNKEKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 121
QY 592 YTGELTNVNNLLKNSFNQNTFLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 651
|||||
Db 122 YTGELTNVNNLLKNSFNQNTFLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 181

QY 652 EKVGKVGEGVGNIANFELDPQYLPQGTQKTYTIASKDYPEVSDGTFTVPTSLAYKMAS 711
|||||
Db 182 EKVGKVGEGVGNIANFELDPQYLPQGTQKTYTIASKDYPEVSDGTFTVPTSLAYKMAS 241
QY 712 QTIETYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNAYLENNYKVGGEIKLPIPKLNQ 771
|||||
Db 242 QTIETYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEFGNAYLENNYKVGGEIKLPIPKLNQ 301
QY 772 GTTTRTAGNKIPVTFMANAYLDNOSTYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLD 831
|||||
Db 302 GTTTRTAGNKIPVTFMANAYLDNOSTYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLD 361
QY 832 EKVEEPTSKVEKEKLSNLTGNSNLTLEEVPTDVPQVKAFABSYGKMLNVLNFM 891
|||||
Db 362 EKVEEPTSKVEKEKLSNLTGNSNLTLEEVPTDVPQVKAFABSYGKMLNVLNFM 421
QY 892 DGTIELYLPSEGEVKKNNMADFTGEAPOGNGENKPSSENGKYSTGTVENQPTENKPADSLPE 951
|||||
Db 422 DGTIELYLPSEGEVKKNNMADFTGEAPOGNGENKPSSENGKYSTGTVENQPTENKPADSLPE 481
QY 952 APNEKPKPNSNDGMLNPEGNVGSDPMLDPALEAPAVDPVQEKLEKTASYGLGLDS 1011
|||||
Db 482 APNEKPKPNSNDGMLNPEGNVGSDPMLDPALEAPAVDPVQEKLEKTASYGLGLDS 541
QY 1012 VIFNMGTIELRLPSGEVKKNNLSDFIA 1039
|||||
Db 542 VIFNMGTIELRLPSGEVKKNNLSDFIA 569

RESULT 19

US-09-884-465A-255
; Sequence 255, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 255
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-255

Query Match 54.7%; Score 568; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLDDKKEEKTAGIMKQYG 531
|||||
Db 2 DLTEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLDDKKEEKTAGIMKQYG 61
QY 532 VKRESIVVNKEKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 591
|||||
Db 62 VKRESIVVNKEKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGNKV 121
QY 592 YTGELTNVNNLLKNSFNQNTFLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 651
|||||
Db 122 YTGELTNVNNLLKNSFNQNTFLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 181
QY 652 EKVGKVGEGVGNIANFELDPQYLPQGTQKTYTIASKDYPEVSDGTFTVPTSLAYKMAS 711

Db 182 EKVSKEGEGVGNIANFELDQYLPQGTFRYITASKDYPEVSDGTFTVPTSLAYKMAS 241
Qy 712 OTIEYFPHAGDTYLRVNPQFAVPKGTDLAVRVDFEFGHGNAYLENNYKVGKIKLPKLNQ 771
Db 242 QTIFYPFHAGDTYLRVNPQFAVPKGTDLAVRVDFEFGHGNAYLENNYKVGKIKLPKLNQ 301
Qy 772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOFKRNKAQENSKLD 831
Db 302 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOFKRNKAQENSKLD 361
Qy 832 EKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTVDPQVEKVAFAESYGMKLENVLFNM 891
Db 362 EKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTVDPQVEKVAFAESYGMKLENVLFNM 421
Qy 892 DGTIELYLPSPGEVIRKKNMADFTGEAPOGNGENKPSKNGKSTGTVENOPTENKPADSLPE 951
Db 422 DGTIELYLPSPGEVIRKKNMADFTGEAPOGNGENKPSKNGKSTGTVENOPTENKPADSLPE 481
Qy 952 APNEKPVKPNSTDNMGNPEGNVSDPMLDPALEEAFAVDPVQVEKLEKFTASYGLGLDS 1011
Db 482 APNEKPVKPNSTDNMGNPEGNVSDPMLDPALEEAFAVDPVQVEKLEKFTASYGLGLDS 541
Qy 1012 VIFNMDGTIELRLPSGEVIRKKNLSDFIA 1039
Db 542 VIFNMDGTIELRLPSGEVIRKKNLSDFIA 569
RESULT 20
US-09-884-465A-384
; Sequence 384, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 384
; LENGTH: 913
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC.FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: xaa = Methionine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (344)..(344)
; OTHER INFORMATION: xaa = Glycine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (345)..(345)
; OTHER INFORMATION: xaa = Proline or nothing
US-09-884-465A-384
Query Match 54.7%; Score 568; DB 9; Length 913;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 472 DLTEEQIKAAQKHLEEVKTSNGLSSSHQDYPGNAKEMKDLDKKIEKIAKIMKQYG 531
Db 346 DLTEEQIKAAQKHLEEVKTSNGLSSSHQDYPGNAKEMKDLDKKIEKIAKIMKQYG 405
Qy 532 VKRESIVVNKKNALIIYPHGDHHDPIDEHKPVGIGHSHSNVELFKPEGVAKKEGKV 591

Db 406 VKRESIVVNKKNALIIYPHGDHHDPIDEHKPVGIGHSHSNVELFKPEGVAKKEGKV 465
Qy 592 YTGEELTNVNLKNTFNQNFLLANGOKRVSFSPPELEKKGITNMLVKLITPDGKVL 651
Db 466 YTGEELTNVNLKNTFNQNFLLANGOKRVSFSPPELEKKGITNMLVKLITPDGKVL 525
Qy 652 EKVSKEGEGVGNIANFELDQYLPQGTFRYITASKDYPEVSDGTFTVPTSLAYKMAS 711
Db 526 EKVSKEGEGVGNIANFELDQYLPQGTFRYITASKDYPEVSDGTFTVPTSLAYKMAS 585
Qy 712 QTIFYPFHAGDTYLRVNPQFAVPKGTDLAVRVDFEFGHGNAYLENNYKVGKIKLPKLNQ 771
Db 586 QTIFYPFHAGDTYLRVNPQFAVPKGTDLAVRVDFEFGHGNAYLENNYKVGKIKLPKLNQ 645
Qy 772 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOFKRNKAQENSKLD 831
Db 646 GTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOFKRNKAQENSKLD 705
Qy 832 EKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTVDPQVEKVAFAESYGMKLENVLFNM 891
Db 706 EKVEEPTSEKVEKEKLSSETGNSNSTLEEVPTVDPQVEKVAFAESYGMKLENVLFNM 765
Qy 892 DGTIELYLPSPGEVIRKKNMADFTGEAPOGNGENKPSKNGKSTGTVENOPTENKPADSLPE 951
Db 766 DGTIELYLPSPGEVIRKKNMADFTGEAPOGNGENKPSKNGKSTGTVENOPTENKPADSLPE 825
Qy 952 APNEKPVKPNSTDNMGNPEGNVSDPMLDPALEEAFAVDPVQVEKLEKFTASYGLGLDS 1011
Db 826 APNEKPVKPNSTDNMGNPEGNVSDPMLDPALEEAFAVDPVQVEKLEKFTASYGLGLDS 885
Qy 1012 VIFNMDGTIELRLPSGEVIRKKNLSDFIA 1039
Db 886 VIFNMDGTIELRLPSGEVIRKKNLSDFIA 913
RESULT 21
US-09-884-465A-376
; Sequence 376, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 376
; LENGTH: 999
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC.FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: xaa = Methionine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (430)..(430)
; OTHER INFORMATION: xaa = Glycine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (431)..(431)
; OTHER INFORMATION: xaa = Proline or nothing
US-09-884-465A-376
Query Match 54.7%; Score 568; DB 9; Length 999;

Best Local Similarity 100.0%; Pred. No. 0:			
Matches	568; Conservative	0; Mismatches	0; Indels
Qy	472	DLTEEQIAKAQKHLVEEYKTSHGNGLSLSHQEQYPCNAKEMKDLDKKEIKTAGIMKQVG	531
Db	432	DLTEEQIAKAQKHLVEEYKTSHGNGLSLSHQEQYPCNAKEMKDLDKKEIKTAGIMKQVG	491
Qy	532	VKRSTVYVNEKNAIITYPHGDHHHADPIDHKPVGTGHSHSNVELFKPPEGVAKKEGKV	591
Db	492	VKRSTVYVNEKNAIITYPHGDHHHADPIDHKPVGTGHSHSNVELFKPPEGVAKKEGKV	551
Qy	592	YTGELINVVNLLKNSFFNQNTFLANGOKRVSFSPPELEKLGINMLVKLITPDGKVL	651
Db	552	YTGELINVVNLLKNSFFNQNTFLANGOKRVSFSPPELEKLGINMLVKLITPDGKVL	611
Qy	652	EKVGSKVFGGVGNIANFELDQBYLPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMAS	711
Db	612	EKVGSKVFGGVGNIANFELDQBYLPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMAS	671
Qy	712	QTIFYFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGEIKLPTPKLQ	771
Db	672	QTIFYFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGEIKLPTPKLQ	731
Qy	772	GTTTATGANKIPVTFMANAYIDNOSTYIVVEPILKEKNQTDKPSILPQFKRNKAQENSKLD	831
Db	732	GTTTATGANKIPVTFMANAYIDNOSTYIVVEPILKEKNQTDKPSILPQFKRNKAQENSKLD	791
Qy	832	EKVEEPTSEKVEKEKLESTGNTSNTLEEVPDQVEKVAFAESYCMKLENNVFNK	891
Db	792	EKVEEPTSEKVEKEKLESTGNTSNTLEEVPDQVEKVAFAESYCMKLENNVFNK	851
Qy	892	DGTTIELYLPSEVITKKNMADFTGEAPQNGENKPNSEKGVSTGTVENQPTENKPADSLPE	951
Db	852	DGTTIELYLPSEVITKKNMADFTGEAPQNGENKPNSEKGVSTGTVENQPTENKPADSLPE	911
Qy	952	APNEKPKVPKNSTNDGMLNPEGNVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGDS	1011
Db	912	APNEKPKVPKNSTNDGMLNPEGNVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGDS	971
Qy	1012	VIFNMDGTIELRLPSGEVIKKNLSDFIA	1039
Db	972	VIFNMDGTIELRLPSGEVIKKNLSDFIA	999

```

RESULT 22
US/9-884-465A-377
; Sequence 377, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 377
; LENGTH: 999
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE

```

```

; LOCATION: (570)...(570)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC FEATURE
; LOCATION: (571)...(571)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-377

Query Match          54.7%; Score 568; DB 9; Length 999;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 472 DLTBEQKAAQKHLEEVKTSINGLSDLSLSHQDYPGNAKEMKDLDKKIEKIAIMKQYG 531
      |||||
Db 2 DLTBEQKAAQKHLEEVKTSINGLSDLSLSHQDYPGNAKEMKDLDKKIEKIAIMKQYG 61

Qy 532 VKRESIVVNEKKNAIYYPHGSHHADPDEHKPVGICGSHSNVELFPEBGVAKKEGNKV 591
      |||||
Db 62 VKRESIVVNEKKNAIYYPHGSHHADPDEHKPVGICGSHSNVELFPEBGVAKKEGNKV 121

Qy 592 YTGELTNVNVLLKNSTFNNOFTLANGQRKVSFSPPELEKLGINMLVKLITPDGKVL 651
      |||||
Db 122 YTGELTNVNVLLKNSTFNNOFTLANGQRKVSFSPPELEKLGINMLVKLITPDGKVL 181

Qy 652 EKVGSKVGEVGNGIANFELDQYLPQGTKYTIASKDXPEVSDGTFVPTSLAYKMAS 711
      |||||
Db 182 EKVGSKVGEVGNGIANFELDQYLPQGTKYTIASKDXPEVSDGTFVPTSLAYKMAS 241

Qy 712 QTITPYFHAGTYLRLVAPQFAVPKGTDALRVDFEFGHNAVLENNYKVGIKLPIPKLNQ 771
      |||||
Db 242 QTITPYFHAGTYLRLVAPQFAVPKGTDALRVDFEFGHNAVLENNYKVGIKLPIPKLNQ 301

Qy 772 GTTTAGNKIPVTMANAYLDNQSTYIIVEPILKEKQTDKPSLLPOFKNKAQENSKLD 831
      |||||
Db 302 GTTTAGNKIPVTMANAYLDNQSTYIIVEPILKEKQTDKPSLLPOFKNKAQENSKLD 361

Qy 832 EKVEEPTKSEKVEKEKLSGTGNSNSTLEEVPTDPVQEKVAKFAESYGMKLENVLFM 891
      |||||
Db 362 EKVEEPTKSEKVEKEKLSGTGNSNSTLEEVPTDPVQEKVAKFAESYGMKLENVLFM 421

Qy 892 DGTIELYLPSEVTKKNADFTEGAPCGNGENKPSSENGKYSTGTVENQPTENKPADSLPE 951
      |||||
Db 422 DGTIELYLPSEVTKKNADFTEGAPCGNGENKPSSENGKYSTGTVENQPTENKPADSLPE 481

Qy 952 APNEKPKVPENSTONGMLNPEGNVGSDDPMLDPALEEAAPVDVQEKLEKFTASYGLGLDS 1011
      |||||
Db 482 APNEKPKVPENSTONGMLNPEGNVGSDDPMLDPALEEAAPVDVQEKLEKFTASYGLGLDS 541

Qy 1012 VIFNMDDGTIELRLPSGEVIRKKNLSDFTA 1039
      |||||
Db 542 VIFNMDDGTIELRLPSGEVIRKKNLSDFTA 569

```

```

RESULT 23
US-09-884-465A-383
; Sequence 383, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent version 3.1
; SEQ ID NO 383
; LENGTH: 1126

```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (557)..(557)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: Xaa = Proline or nothing
; US-09-884-465A-383

Query Match      54.7%: Score 568; DB 9; Length 1126;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEQIKAAQKHLVEVKTSHNGDLSLSSHEQDYPGNAKEMKDLDDKKIEKTAGIMKQYG 531
DB 559 DLTEQIKAAQKHLVEVKTSHNGDLSLSSHEQDYPGNAKEMKDLDDKKIEKTAGIMKQYG 618
QY 532 VKRESIVNKEKNAIYPHGDHHDHAPDIDHKPVGIGHSHSNYELFKPEEGVAKKEGKV 591
DB 619 VKRESIVNKEKNAIYPHGDHHDHAPDIDHKPVGIGHSHSNYELFKPEEGVAKKEGKV 678
QY 592 YTGEELTNVNLLKKNSTFNQNFLLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 651
DB 679 YTGEELTNVNLLKKNSTFNQNFLLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 738
QY 652 EKVSQKVFGEVGNIANFELDPYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMAS 711
DB 739 EKVSQKVFGEVGNIANFELDPYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMAS 798
QY 712 QTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHNAYLENNYKVGEIKLPIPKLNQ 771
DB 799 QTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHNAYLENNYKVGEIKLPIPKLNQ 858
QY 772 GTTRTAGNKIPVTFMANAYLDNOSTYIIVEVPILEKENOTDKPSIIPQKRNKAQENSKLD 831
DB 859 GTTRTAGNKIPVTFMANAYLDNOSTYIIVEVPILEKENOTDKPSIIPQKRNKAQENSKLD 918
QY 832 EKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 891
DB 919 EKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 978
QY 892 DGTIELPSGVEIKKNMADFTGEAPQNGENKPSGKSTGTGTVENOPTENKPADSLPE 951
DB 979 DGTIELPSGVEIKKNMADFTGEAPQNGENKPSGKSTGTGTVENOPTENKPADSLPE 1038
QY 952 APNEKVPKPNSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
DB 1039 APNEKVPKPNSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1098
QY 1012 VIFNMDGTIELRPSGEVKKNLSDFIA 1039
DB 1099 VIFNMDGTIELRPSGEVKKNLSDFIA 1126

RESULT 24
US-09-884-465A-380
; Sequence 380, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Saire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
```

```
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 380
; LENGTH: 1139
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (570)..(570)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (571)..(571)
; OTHER INFORMATION: Xaa = Proline or nothing
; US-09-884-465A-380

Query Match      54.7%: Score 568; DB 9; Length 1139;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 DLTEQIKAAQKHLVEVKTSHNGDLSLSSHEQDYPGNAKEMKDLDDKKIEKTAGIMKQYG 531
DB 572 DLTEQIKAAQKHLVEVKTSHNGDLSLSSHEQDYPGNAKEMKDLDDKKIEKTAGIMKQYG 631
QY 532 VKRESIVNKEKNAIYPHGDHHDHAPDIDHKPVGIGHSHSNYELFKPEEGVAKKEGKV 591
DB 632 VKRESIVNKEKNAIYPHGDHHDHAPDIDHKPVGIGHSHSNYELFKPEEGVAKKEGKV 691
QY 592 YTGEELTNVNLLKKNSTFNQNFLLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 651
DB 692 YTGEELTNVNLLKKNSTFNQNFLLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVL 751
QY 652 EKVSQKVFGEVGNIANFELDPYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMAS 711
DB 752 EKVSQKVFGEVGNIANFELDPYLPQGTFTYTIASKDYPEVSYDGTFTVPTSLAYKMAS 811
QY 712 QTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHNAYLENNYKVGEIKLPIPKLNQ 771
DB 812 QTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHNAYLENNYKVGEIKLPIPKLNQ 871
QY 772 GTTRTAGNKIPVTFMANAYLDNOSTYIIVEVPILEKENOTDKPSIIPQKRNKAQENSKLD 831
DB 872 GTTRTAGNKIPVTFMANAYLDNOSTYIIVEVPILEKENOTDKPSIIPQKRNKAQENSKLD 931
QY 832 EKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 891
DB 932 EKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTDVPQEKVAKFAESYGMKLENVLFNM 991
QY 892 DGTIELPSGVEIKKNMADFTGEAPQNGENKPSGKSTGTGTVENOPTENKPADSLPE 951
DB 992 DGTIELPSGVEIKKNMADFTGEAPQNGENKPSGKSTGTGTVENOPTENKPADSLPE 1051
QY 952 APNEKVPKPNSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1011
DB 1052 APNEKVPKPNSTDNGLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDS 1111
QY 1012 VIFNMDGTIELRPSGEVKKNLSDFIA 1039
DB 1112 VIFNMDGTIELRPSGEVKKNLSDFIA 1139

RESULT 25
US-09-884-465A-354
; Sequence 354, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
```


; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 354
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-354

Query Match 51.1%; Score 531; DB 9; Length 633;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 631; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 408 QIGOPTLPNNSLATPSPSLPINCTSHKHEEDCYGFDANRIAEDESFGVMSHGDNHY 467
Db 2 QIGOPTLPNNSLATPSPSLPINCTSHKHEEDCYGFDANRIAEDESFGVMSHGDNHY 61
QY 468 FFKKDLTEEQIAAKQKLEEVKTSNGLDLSLSHSEQDPYGNAKEMKDKKIEKIAIGM 527
Db 62 FFKKDLTEEQIAAKQKLEEVKTSNGLDLSLSHSEQDPYGNAKEMKDKKIEKIAIGM 121
QY 528 KOYGKRESIVVNKKNAILYPHGDHHDPIDEHKPVGIGHSHSNVELPKPEGVAKKE 587
Db 122 KOYGKRESIVVNKKNAILYPHGDHHDPIDEHKPVGIGHSHSNVELPKPEGVAKKE 181
QY 588 GNKYTGEBELTNVYNLLKNSTFNQNETLANGQKRVSFSEPPPELEKLGINMLVKLITPD 647
Db 182 GNKYTGEBELTNVYNLLKNSTFNQNETLANGQKRVSFSEPPPELEKLGINMLVKLITPD 241
QY 648 GKVLEKSGKVGEGVGNIANFELDQYLPQOTFKYTIASKDYPEVSYDGTFTVPTSLAY 707
Db 242 GKVLEKSGKVGEGVGNIANFELDQYLPQOTFKYTIASKDYPEVSYDGTFTVPTSLAY 301
QY 708 KMASQTIFFYPFHAGDTYLVRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 767
Db 302 KMASQTIFFYPFHAGDTYLVRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPI 361
QY 768 KLNQGTTRTAGNKIPVTFMANAYLDNOSTYIVVEVPILEKENQTDKPSILPQFKRNKAQEN 827
Db 362 KLNQGTTRTAGNKIPVTFMANAYLDNOSTYIVVEVPILEKENQTDKPSILPQFKRNKAQEN 421
QY 828 SKLDEKVEEPTSEKVEKEKLSNSTSTLEEVPTDVPQEKVAKFAESYGMKLENY 887
Db 422 SKLDEKVEEPTSEKVEKEKLSNSTSTLEEVPTDVPQEKVAKFAESYGMKLENY 481
QY 888 LFNMDGTIELYPSGEVIKKNMADFTGEAPQNGENKPSNGKSTGTVENQPTENKPAD 947
Db 482 LFNMDGTIELYPSGEVIKKNMADFTGEAPQNGENKPSNGKSTGTVENQPTENKPAD 541
QY 948 SLPEAPNEKVPKPNSTNDGMLNPEGVNSDPMLDPALEAPADVPQEKLEKFTASYGL 1007
Db 542 SLPEAPNEKVPKPNSTNDGMLNPEGVNSDPMLDPALEAPADVPQEKLEKFTASYGL 601
QY 1008 GLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1039
Db 602 GLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 633

RESULT 26
US-09-884-465A-352

; Sequence 352, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 352
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-352

Query Match 47.1%; Score 489; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 551 GDHEHADPIDDEHKPVGIGHSHSNVELFKPEEGVAKKEGKNVYTGELTNVYNLLKNSTFN 610
Db 145 GDHEHADPIDDEHKPVGIGHSHSNVELFKPEEGVAKKEGKNVYTGELTNVYNLLKNSTFN 204
QY 611 NQNETLANGQKRVSFSEPPPELEKLGINMLVKLITPDGKVLKSGKVGEGVGNIANEE 670
Db 205 NQNETLANGQKRVSFSEPPPELEKLGINMLVKLITPDGKVLKSGKVGEGVGNIANEE 264
QY 671 LDQPYLPQOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLVRVNPQ 730
Db 265 LDQPYLPQOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLVRVNPQ 324
QY 731 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 790
Db 325 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 384
QY 791 LDNQSYIVVEVPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 850
Db 385 LDNQSYIVVEVPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 444
QY 851 TGNSTNSLTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMDGTIELYPSGEVIKKNMA 910
Db 445 TGNSTNSLTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMDGTIELYPSGEVIKKNMA 504
QY 911 DFTGEAPQNGENKPSNGKSTGTVENQPTENKPADSLPEAPNEKVPKPNSTNDGMLN 970
Db 505 DFTGEAPQNGENKPSNGKSTGTVENQPTENKPADSLPEAPNEKVPKPNSTNDGMLN 564
QY 971 PEGNVGSDPMLDPALEAPADVPQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVI 1030
Db 565 PEGNVGSDPMLDPALEAPADVPQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVI 624
QY 1031 KKNLSDFIA 1039
Db 625 KKNLSDFIA 633

RESULT 27
US-09-884-465A-357
; Sequence 357, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee

APPLICANT: Brodeur, Bernard.
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 357
LENGTH: 633
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-357

Query Match 47.1%; Score 489; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 551 GDHHADPIDEHKPVGIGHSHSNVELEFKPEGVAKKEGKNVYTGELTNVYNLLKNSTFN 610
DB 145 GDHHADPIDEHKPVGIGHSHSNVELEFKPEGVAKKEGKNVYTGELTNVYNLLKNSTFN 204
QY 611 NQFTLANGOKRVSFSPPELEKLGINMLVKLITPDGKYLEKVSCKVFGVGNIANFE 670
DB 205 NQFTLANGOKRVSFSPPELEKLGINMLVKLITPDGKYLEKVSCKVFGVGNIANFE 264
QY 671 LDQYLPQGTFTKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 730
DB 265 LDQYLPQGTFTKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 324
QY 731 FAVPKGTDALVRVDEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 790
DB 325 FAVPKGTDALVRVDEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 384
QY 791 LDQSTYIYVEVPILEKENQDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 850
DB 385 LDQSTYIYVEVPILEKENQDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 444
QY 851 TGNSTSNSTLEEVPVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSSGEVIKKNMA 910
DB 445 TGNSTSNSTLEEVPVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSSGEVIKKNMA 504
QY 911 DFTGEAPQNGENKPSNGKVSCTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLN 970
DB 505 DFTGEAPQNGENKPSNGKVSCTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLN 564
QY 971 PEGNVGSDPMDLPALEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVI 1030
DB 565 PEGNVGSDPMDLPALEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVI 624
QY 1031 KKNLSDFIA 1039
DB 625 KKNLSDFIA 633

RESULT 28
US-09-884-465A-358
Sequence 358, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biotech, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens

FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 358
LENGTH: 633
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-358

Query Match 47.1%; Score 489; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 551 GDHHADPIDEHKPVGIGHSHSNVELEFKPEGVAKKEGKNVYTGELTNVYNLLKNSTFN 610
DB 145 GDHHADPIDEHKPVGIGHSHSNVELEFKPEGVAKKEGKNVYTGELTNVYNLLKNSTFN 204
QY 611 NQFTLANGOKRVSFSPPELEKLGINMLVKLITPDGKYLEKVSCKVFGVGNIANFE 670
DB 205 NQFTLANGOKRVSFSPPELEKLGINMLVKLITPDGKYLEKVSCKVFGVGNIANFE 264
QY 671 LDQYLPQGTFTKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 730
DB 265 LDQYLPQGTFTKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 324
QY 731 FAVPKGTDALVRVDEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 790
DB 325 FAVPKGTDALVRVDEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 384
QY 791 LDQSTYIYVEVPILEKENQDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 850
DB 385 LDQSTYIYVEVPILEKENQDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 444
QY 851 TGNSTSNSTLEEVPVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSSGEVIKKNMA 910
DB 445 TGNSTSNSTLEEVPVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSSGEVIKKNMA 504
QY 911 DFTGEAPQNGENKPSNGKVSCTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLN 970
DB 505 DFTGEAPQNGENKPSNGKVSCTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLN 564
QY 971 PEGNVGSDPMDLPALEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVI 1030
DB 565 PEGNVGSDPMDLPALEAPAVDPVOEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVI 624
QY 1031 KKNLSDFIA 1039
DB 625 KKNLSDFIA 633

RESULT 29
US-09-884-465A-359
Sequence 359, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biotech, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20

```
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 359
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-359

Query Match          47.1%; Score 489; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHADPIDEHKPVGIGHSHSNVELFKPEGVAKEGKNKYVTGPELTNNVNLKNSTEN 610
    |||||
DB 145 GDHHADPIDEHKPVGIGHSHSNVELFKPEGVAKEGKNKYVTGPELTNNVNLKNSTEN 204
    |||||
QY 611 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKESGKVFGEVGNIANFE 670
    |||||
DB 205 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKESGKVFGEVGNIANFE 264
    |||||
QY 671 LDQPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ 730
    |||||
DB 265 LDQPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ 324
    |||||
QY 731 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 790
    |||||
DB 325 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 384
    |||||
QY 791 LDNQSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPKTSKVEKEKLE 850
    |||||
DB 385 LDNQSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPKTSKVEKEKLE 444
    |||||
QY 851 TGNSTNSLTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMA 910
    |||||
DB 445 TGNSTNSLTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMA 504
    |||||
QY 911 DFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLN 970
    |||||
DB 505 DFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLN 564
    |||||
QY 971 PEGNVGSDPMLDPALEAPADVPQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
    |||||
DB 565 PEGNVGSDPMLDPALEAPADVPQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 624
    |||||
QY 1031 KKNLSDFIA 1039
    |||||
DB 625 KKNLSDFIA 633
    |||||

RESULT 30
US-09-884-465A-360
; Sequence 360, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; LENGTH: 633
; TYPE: PRT
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-360

Query Match          47.1%; Score 489; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHADPIDEHKPVGIGHSHSNVELFKPEGVAKEGKNKYVTGPELTNNVNLKNSTEN 610
    |||||
DB 145 GDHHADPIDEHKPVGIGHSHSNVELFKPEGVAKEGKNKYVTGPELTNNVNLKNSTEN 204
    |||||
QY 611 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKESGKVFGEVGNIANFE 670
    |||||
DB 205 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKESGKVFGEVGNIANFE 264
    |||||
QY 671 LDQPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ 730
    |||||
DB 265 LDQPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ 324
    |||||
QY 731 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 790
    |||||
DB 325 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAY 384
    |||||
QY 791 LDNQSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPKTSKVEKEKLE 850
    |||||
DB 385 LDNQSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPKTSKVEKEKLE 444
    |||||
QY 851 TGNSTNSLTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMA 910
    |||||
DB 445 TGNSTNSLTLEEVPTDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMA 504
    |||||
QY 911 DFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLN 970
    |||||
DB 505 DFTGEAPQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLN 564
    |||||
QY 971 PEGNVGSDPMLDPALEAPADVPQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
    |||||
DB 565 PEGNVGSDPMLDPALEAPADVPQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 624
    |||||
QY 1031 KKNLSDFIA 1039
    |||||
DB 625 KKNLSDFIA 633
    |||||

RESULT 31
US-09-884-465A-333
; Sequence 333, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 333
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-333
```

```
Query Match 47.1%; Score 489; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKEGKNKYTGELTNVNLKNSTFN 610
DB 145 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKEGKNKYTGELTNVNLKNSTFN 204
QY 611 NQFTLANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKSVGKVGEGVGNIANFE 670
DB 205 NQFTLANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKSVGKVGEGVGNIANFE 264
QY 671 LDQYLPFGQTFKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 730
DB 265 LDQYLPFGQTFKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 324
QY 731 FAVPKGTDALVRVDFEHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAY 790
DB 325 FAVPKGTDALVRVDFEHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAY 384
QY 791 LDNOSTYIIVEPILKEKNTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 850
DB 385 LDNOSTYIIVEPILKEKNTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 444
QY 851 TGNSTNSLIEEVPVDPVQEKVAKFAESYGMKLENVLFNMDGTIELRPSGEVI 1030
DB 445 TGNSTNSLIEEVPVDPVQEKVAKFAESYGMKLENVLFNMDGTIELRPSGEVI 504
QY 911 DFTGEAPOGNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKPKPENSTDMGLN 970
DB 505 DFTGEAPOGNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKPKPENSTDMGLN 564
QY 971 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRPSGEVI 1030
DB 565 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRPSGEVI 624
QY 1031 KKNLSDFIA 1039
DB 625 KKNLSDFIA 633

RESULT 32
US-09-884-465A-334
; Sequence 334, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884.465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 334
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-334

Query Match 47.1%; Score 489; DB 9; Length 900;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKEGKNKYTGELTNVNLKNSTFN 610
DB 418 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKEGKNKYTGELTNVNLKNSTFN 477
QY 611 NQFTLANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKSVGKVGEGVGNIANFE 670
```

```
DB 412 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKEGKNKYTGELTNVNLKNSTFN 471
QY 611 NQFTLANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKSVGKVGEGVGNIANFE 670
DB 472 NQFTLANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKSVGKVGEGVGNIANFE 531
QY 671 LDQYLPFGQTFKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 730
DB 532 LDQYLPFGQTFKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQ 591
QY 731 FAVPKGTDALVRVDFEHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAY 790
DB 592 FAVPKGTDALVRVDFEHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAY 651
QY 791 LDNOSTYIIVEPILKEKNTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 850
DB 652 LDNOSTYIIVEPILKEKNTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLS 711
QY 851 TGNSTNSLIEEVPVDPVQEKVAKFAESYGMKLENVLFNMDGTIELRPSGEVI 910
DB 712 TGNSTNSLIEEVPVDPVQEKVAKFAESYGMKLENVLFNMDGTIELRPSGEVI 771
QY 911 DFTGEAPOGNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKPKPENSTDMGLN 970
DB 772 DFTGEAPOGNGENKPSGKSTGTVENQPTENKPADSLPEAPNEKPKPENSTDMGLN 831
QY 971 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRPSGEVI 1030
DB 832 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRPSGEVI 891
QY 1031 KKNLSDFIA 1039
DB 892 KKNLSDFIA 900

RESULT 33
US-09-884-465A-332
; Sequence 332, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884.465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 332
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-332

Query Match 47.1%; Score 489; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKEGKNKYTGELTNVNLKNSTFN 610
DB 418 GDHHDADPIDEHKPVGIGHSHSNYELFKPEGVAKEGKNKYTGELTNVNLKNSTFN 477
QY 611 NQFTLANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKSVGKVGEGVGNIANFE 670
```

Db 478 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKSGKVFGEVGVGNIANFE 537
Qy 671 LDOPYLPQGTFTKTYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPPHAGDTYLVRNPQ 730
Db 538 LDOPYLPQGTFTKTYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPPHAGDTYLVRNPQ 597
Qy 731 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 790
Db 598 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 657
Qy 791 LDONQSTYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSE 850
Db 658 LDONQSTYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSE 717
Qy 851 TGNSTNSNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKNMA 910
Db 718 TGNSTNSNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKNMA 777
Qy 911 DFTGEAPQNGENKPSKNGKVTGTVENQPTENKPADSLPEAPNEKPKVPKPNSTNDGMLN 970
Db 778 DFTGEAPQNGENKPSKNGKVTGTVENQPTENKPADSLPEAPNEKPKVPKPNSTNDGMLN 837
Qy 971 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSCGEVI 1030
Db 838 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSCGEVI 897
Qy 1031 KKNLSDFIA 1039
Db 898 KKNLSDFIA 906
RESULT 34
US-09-884-465A-369
; Sequence 369, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 369
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-369
Query Match 47.1%; Score 489; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 551 GDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGEELTNVNLLKNSTFN 610
Db 145 GDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGEELTNVNLLKNSTFN 204
Qy 611 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKSGKVFGEVGVGNIANFE 670
Db 205 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKSGKVFGEVGVGNIANFE 264
Qy 671 LDOPYLPQGTFTKTYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPPHAGDTYLVRNPQ 730
Db 265 LDOPYLPQGTFTKTYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPPHAGDTYLVRNPQ 324

Qy 731 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 790
Db 325 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 384
Qy 791 LDONQSTYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSE 850
Db 385 LDONQSTYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSE 444
Qy 851 TGNSTNSNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKNMA 910
Db 445 TGNSTNSNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSCGEVIKKNMA 504
Qy 911 DFTGEAPQNGENKPSKNGKVTGTVENQPTENKPADSLPEAPNEKPKVPKPNSTNDGMLN 970
Db 505 DFTGEAPQNGENKPSKNGKVTGTVENQPTENKPADSLPEAPNEKPKVPKPNSTNDGMLN 564
Qy 971 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSCGEVI 1030
Db 565 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSCGEVI 624
Qy 1031 KKNLSDFIA 1039
Db 625 KKNLSDFIA 633
RESULT 35
US-09-884-465A-370
; Sequence 370, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 370
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-370
Query Match 47.1%; Score 489; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 551 GDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGEELTNVNLLKNSTFN 610
Db 418 GDHHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGEELTNVNLLKNSTFN 477
Qy 611 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKSGKVFGEVGVGNIANFE 670
Db 478 NONFTLANGQKRVSFSPPELEKKGILNMLVKLITPDGKVLKSGKVFGEVGVGNIANFE 537
Qy 671 LDOPYLPQGTFTKTYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPPHAGDTYLVRNPQ 730
Db 538 LDOPYLPQGTFTKTYTIASKDYPEVSYDGTFTVPTSLAYKASOTIFYPPHAGDTYLVRNPQ 597
Qy 731 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 790
Db 598 FAVPKGTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTRTAGNKIPVTFMANAY 657

```
QY 791 LDNQSTYIIEVPILEKRENOTDKPSILPQFKRNKAQNSKLDKEKVEEPTSEKVEKEKLS 850
Db 658 LDNQSTYIIEVPILEKRENOTDKPSILPQFKRNKAQNSKLDKEKVEEPTSEKVEKEKLS 717
QY 851 TGNSTNSLTLEEYPTVDVPOEKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNNA 910
Db 718 TGNSTNSLTLEEYPTVDVPOEKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNNA 777
QY 911 DFTGEAPOGNGENKPSNGKSVSTGTVENOPTENKPADSLPEAPNEKPKVPENSTNGMLN 970
Db 778 DFTGEAPOGNGENKPSNGKSVSTGTVENOPTENKPADSLPEAPNEKPKVPENSTNGMLN 837
QY 971 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
Db 838 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 897
QY 1031 KKNLSDFIA 1039
Db 898 KKNLSDFIA 906

RESULT 36
US-09-884-465A-373
; Sequence 373, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Ouellet, Catherine
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Ouellet, Catherine
; APPLICANT: Hamel, Josee
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 373
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-373

Query Match 47.1%; Score 489; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 GDHHDADTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVNVLLKNSTFN 610
Db 145 GDHHDADTDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVNVLLKNSTFN 204
QY 611 NONFTLANGOKRVSFSPPELEKLGINMLVLTIPDGKVLKESKVGEGVGNIANFE 670
Db 205 NONFTLANGOKRVSFSPPELEKLGINMLVLTIPDGKVLKESKVGEGVGNIANFE 264
QY 671 LDQYLPQGTFTFYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ 730
Db 265 LDQYLPQGTFTFYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ 324
QY 731 FAVPKGTDLVRVDFDEHGNAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTMANAY 790
Db 325 FAVPKGTDLVRVDFDEHGNAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTMANAY 384
QY 791 LDNQSTYIIEVPILEKRENOTDKPSILPQFKRNKAQNSKLDKEKVEEPTSEKVEKEKLS 850
Db 385 LDNQSTYIIEVPILEKRENOTDKPSILPQFKRNKAQNSKLDKEKVEEPTSEKVEKEKLS 444
QY 851 TGNSTNSLTLEEYPTVDVPOEKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNNA 910
```

```
Db 445 TGNSTNSLTLEEYPTVDVPOEKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNMA 504
QY 911 DFTGEAPOGNGENKPSNGKSVSTGTVENOPTENKPADSLPEAPNEKPKVPENSTNGMLN 970
Db 505 DFTGEAPOGNGENKPSNGKSVSTGTVENOPTENKPADSLPEAPNEKPKVPENSTNGMLN 564
QY 971 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 1030
Db 565 PEGNVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVI 624
QY 1031 KKNLSDFIA 1039
Db 625 KKNLSDFIA 633

RESULT 37
US-09-884-465A-356
; Sequence 356, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 356
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-356

Query Match 46.8%; Score 486; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVNVLLKNSTFNQN 613
Db 84 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVNVLLKNSTFNQN 143
QY 614 FTLANGOKRVSFSPPELEKLGINMLVLTIPDGKVLKESKVGEGVGNIANFELDQ 673
Db 144 FTLANGOKRVSFSPPELEKLGINMLVLTIPDGKVLKESKVGEGVGNIANFELDQ 203
QY 674 PYLPGQTFKTYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 733
Db 204 PYLPGQTFKTYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 263
QY 734 PKGTDLVRVDFDEHGNAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTMANAYLDN 793
Db 264 PKGTDLVRVDFDEHGNAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTMANAYLDN 323
QY 794 QSTVIVEVPILEKENQTDKPSILPQFKRNKAQNSKLDKEKVEEPTSEKVEKEKLS 853
Db 324 QSTVIVEVPILEKENQTDKPSILPQFKRNKAQNSKLDKEKVEEPTSEKVEKEKLS 383
QY 854 STNSTLEEYPTVDVPOEKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNMAFFT 913
Db 384 STNSTLEEYPTVDVPOEKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNMAFFT 443
QY 914 GEAPQNGENKPSNGKSVSTGTVENOPTENKPADSLPEAPNEKPKVPENSTNGMLNPEG 973
```

Db 444 GEAQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEG 503

Qy 974 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 1033

Db 504 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 563

Qy 1034 LSDFIA 1039

Db 564 LSDFIA 569

RESULT 38

US-09-884-465A-353

; Sequence 353, Application US/09884465A

; Publication No. US20030077293A1

; GENERAL INFORMATION:

; APPLICANT: Shire Biochem, Inc.

; APPLICANT: Hamel, Josee

; APPLICANT: Brodeur, Bernard

; APPLICANT: Martin, Denis

; APPLICANT: Charland, Nathalie

; APPLICANT: Ouellet, Catherine

; TITLE OF INVENTION: Streptococcus Antigens

; FILE REFERENCE: 055190-0044

; CURRENT APPLICATION NUMBER: US/09/884,465A

; PRIOR FILING DATE: 2001-06-20

; PRIOR FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 384

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 353

; LENGTH: 633

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Unknown Organism

US-09-884-465A-353

Query Match 46.8%; Score 486; DB 9; Length 633;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 554 HHADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKYVTGEELTNNVNLKSTNNQN 613

Db 148 HHADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKYVTGEELTNNVNLKSTNNQN 207

Qy 614 FTLANGQKRVSFSPPELEKKGINMLVKLITPDGKLVLEKSGVFGVGNIANFELDQ 673

Db 208 FTLANGQKRVSFSPPELEKKGINMLVKLITPDGKLVLEKSGVFGVGNIANFELDQ 267

Qy 674 PYLPQGTFTKTYIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 733

Db 268 PYLPQGTFTKTYIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 327

Qy 734 PKGTDALVRVDFEHGNAYLENNYKVGIEIKLPIKLNQGTTRTAGNKIPVTFMANAYLDN 793

Db 328 PKGTDALVRVDFEHGNAYLENNYKVGIEIKLPIKLNQGTTRTAGNKIPVTFMANAYLDN 387

Qy 794 QSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 853

Db 388 QSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 447

Qy 854 STSNSTILEEVPVTPDQVEKVAFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNADFT 913

Db 448 STSNSTILEEVPVTPDQVEKVAFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNADFT 507

Qy 914 GEAQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEG 567

Db 508 GEAQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEG 567

Qy 974 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 1033

Db 568 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 627

Qy 1034 LSDFIA 1039

Db 628 LSDFIA 633

RESULT 39

US-09-884-465A-355

; Sequence 355, Application US/09884465A

; Publication No. US20030077293A1

; GENERAL INFORMATION:

; APPLICANT: Shire Biochem, Inc.

; APPLICANT: Hamel, Josee

; APPLICANT: Brodeur, Bernard

; APPLICANT: Martin, Denis

; APPLICANT: Charland, Nathalie

; APPLICANT: Ouellet, Catherine

; TITLE OF INVENTION: Streptococcus Antigens

; FILE REFERENCE: 055190-0044

; CURRENT APPLICATION NUMBER: US/09/884,465A

; PRIOR FILING DATE: 2001-06-20

; PRIOR FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 384

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 355

; LENGTH: 633

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Unknown Organism

US-09-884-465A-355

Query Match 46.8%; Score 486; DB 9; Length 633;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 554 HHADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKYVTGEELTNNVNLKSTNNQN 613

Db 148 HHADPIDEHKPVGIGHSHSNYELFKPEGVAKKGNKYVTGEELTNNVNLKSTNNQN 207

Qy 614 FTLANGQKRVSFSPPELEKKGINMLVKLITPDGKLVLEKSGVFGVGNIANFELDQ 673

Db 208 FTLANGQKRVSFSPPELEKKGINMLVKLITPDGKLVLEKSGVFGVGNIANFELDQ 267

Qy 674 PYLPQGTFTKTYIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 733

Db 268 PYLPQGTFTKTYIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 327

Qy 734 PKGTDALVRVDFEHGNAYLENNYKVGIEIKLPIKLNQGTTRTAGNKIPVTFMANAYLDN 793

Db 328 PKGTDALVRVDFEHGNAYLENNYKVGIEIKLPIKLNQGTTRTAGNKIPVTFMANAYLDN 387

Qy 794 QSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 853

Db 388 QSTYIVVEPPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 447

Qy 854 STSNSTILEEVPVTPDQVEKVAFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNADFT 913

Db 448 STSNSTILEEVPVTPDQVEKVAFAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNADFT 507

Qy 914 GEAQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEG 567

Db 508 GEAQNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEG 567

Qy 974 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 1033

Db 568 NVGSDPMLDPALEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVIKKN 627

Qy 1034 LSDFIA 1039

Db 628 LSDFIA 633

RESULT 40
US-09-884-465A-363
; Sequence 363, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 363
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-363

Query Match 46.8%; Score 486; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVYVLLKKNSTNNQ 613
DB 148 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVYVLLKKNSTNNQ 207
QY 614 FTLANGOKRVSFSPPELEKKGINMLVLIITPDGKVLKSVKGVGEGVGNIANFELDQ 673
DB 208 FTLANGOKRVSFSPPELEKKGINMLVLIITPDGKVLKSVKGVGEGVGNIANFELDQ 267
QY 674 PYPGQTFYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 733
DB 268 PYPGQTFYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 327
QY 734 PKGTDLVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTAGNKIPVTFMANAYLDN 793
DB 328 PKGTDLVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTAGNKIPVTFMANAYLDN 387
QY 794 QSTYIVVEVPILKENVQDKPSILPQFKRNAQENSKLDKVEEPTSEKVEKEKLSGTGN 853
DB 388 QSTYIVVEVPILKENVQDKPSILPQFKRNAQENSKLDKVEEPTSEKVEKEKLSGTGN 447
QY 854 STNSNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMAFT 913
DB 448 STNSNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMAFT 507
QY 914 GEAPQNGENKPSGKSVGTGVENQPTENKPADSLPEAPNEKPVKPNSTDNGMLNPEG 973
DB 508 GEAPQNGENKPSGKSVGTGVENQPTENKPADSLPEAPNEKPVKPNSTDNGMLNPEG 567
QY 974 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSDVIFNMDGTIELRPSGEVIKKN 1033
DB 568 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSDVIFNMDGTIELRPSGEVIKKN 627
QY 1034 LSDFIA 1039
DB 628 LSDFIA 633

RESULT 41
US-09-884-465A-364
; Sequence 364, Application US/09884465A
; Publication No. US20030077293A1

; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 364
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-364

Query Match 46.8%; Score 486; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVYVLLKKNSTNNQ 613
DB 148 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKKEGKNVYTGELTNVYVLLKKNSTNNQ 207
QY 614 FTLANGOKRVSFSPPELEKKGINMLVLIITPDGKVLKSVKGVGEGVGNIANFELDQ 673
DB 208 FTLANGOKRVSFSPPELEKKGINMLVLIITPDGKVLKSVKGVGEGVGNIANFELDQ 267
QY 674 PYPGQTFYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 733
DB 268 PYPGQTFYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAV 327
QY 734 PKGTDLVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTAGNKIPVTFMANAYLDN 793
DB 328 PKGTDLVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTAGNKIPVTFMANAYLDN 387
QY 794 QSTYIVVEVPILKENVQDKPSILPQFKRNAQENSKLDKVEEPTSEKVEKEKLSGTGN 853
DB 388 QSTYIVVEVPILKENVQDKPSILPQFKRNAQENSKLDKVEEPTSEKVEKEKLSGTGN 447
QY 854 STNSNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMAFT 913
DB 448 STNSNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMAFT 507
QY 914 GEAPQNGENKPSGKSVGTGVENQPTENKPADSLPEAPNEKPVKPNSTDNGMLNPEG 973
DB 508 GEAPQNGENKPSGKSVGTGVENQPTENKPADSLPEAPNEKPVKPNSTDNGMLNPEG 567
QY 974 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSDVIFNMDGTIELRPSGEVIKKN 1033
DB 568 NVGSDPMLDPALEAPAVDPVQEKLEKFTASYGLDSDVIFNMDGTIELRPSGEVIKKN 627
QY 1034 LSDFIA 1039
DB 628 LSDFIA 633

RESULT 42
US-09-884-465A-365
; Sequence 365, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis

APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 365
LENGTH: 633
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-365

Query Match 46.8%; Score 486; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKGEGKVKVTGEEELTNVNNLLKSTFNQNN 613
DB 148 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKGEGKVKVTGEEELTNVNNLLKSTFNQNN 207
QY 614 FTLANGOKRVSPFPPELEKKLGINMLVKLITPDGKVLKGVSKVFGVGVGNIANFELQ 673
DB 208 FTLANGOKRVSPFPPELEKKLGINMLVKLITPDGKVLKGVSKVFGVGVGNIANFELQ 267
QY 674 PYPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 733
DB 268 PYPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 327
QY 734 PKGTDALVRVDFEFGHNAVLENNYKVGGEIKLPIPKLQGTTRTAGNKIPVTFMANAYLDN 793
DB 328 PKGTDALVRVDFEFGHNAVLENNYKVGGEIKLPIPKLQGTTRTAGNKIPVTFMANAYLDN 387
QY 794 QSTYIVVEPILKENDQDKPSILPQFKNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 853
DB 388 QSTYIVVEPILKENDQDKPSILPQFKNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 447
QY 854 STNSNLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNMADFT 913
DB 448 STNSNLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNMADFT 507
QY 914 GEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLNPEG 973
DB 508 GEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLNPEG 567
QY 974 NVGSDPMLDPALPEAPVDPQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVKKN 1033
DB 568 NVGSDPMLDPALPEAPVDPQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVKKN 627

QY 1034 LSDFIA 1039
DB 628 LSDFIA 633

RESULT 43
US-09-884-465A-371
Sequence 371, Application US/09884465A
Publication No. US2003007293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A

CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 371
LENGTH: 906
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-371

Query Match 46.8%; Score 486; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKGEGKVKVTGEEELTNVNNLLKSTFNQNN 613
DB 148 HHADPIDEHKPVGIGHSHSNYELFKPEEGVAKGEGKVKVTGEEELTNVNNLLKSTFNQNN 207
QY 614 FTLANGOKRVSPFPPELEKKLGINMLVKLITPDGKVLKGVSKVFGVGVGNIANFELQ 673
DB 208 FTLANGOKRVSPFPPELEKKLGINMLVKLITPDGKVLKGVSKVFGVGVGNIANFELQ 267
QY 674 PYPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 733
DB 268 PYPGOTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFV 327
QY 734 PKGTDALVRVDFEFGHNAVLENNYKVGGEIKLPIPKLQGTTRTAGNKIPVTFMANAYLDN 793
DB 328 PKGTDALVRVDFEFGHNAVLENNYKVGGEIKLPIPKLQGTTRTAGNKIPVTFMANAYLDN 387
QY 794 QSTYIVVEPILKENDQDKPSILPQFKNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 853
DB 388 QSTYIVVEPILKENDQDKPSILPQFKNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 447
QY 854 STNSNLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNMADFT 913
DB 448 STNSNLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVKKNMADFT 507
QY 914 GEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLNPEG 973
DB 508 GEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLNPEG 567
QY 974 NVGSDPMLDPALPEAPVDPQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVKKN 1033
DB 568 NVGSDPMLDPALPEAPVDPQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSPGEVKKN 627

QY 1034 LSDFIA 1039
DB 628 LSDFIA 633

RESULT 44
US-09-884-465A-372
Sequence 372, Application US/09884465A
Publication No. US2003007293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1

```
; SEQ ID NO 372
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-372

Query Match          46.8%; Score 486; DB 9; Length 906;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 554 HHADPIDEHKPVGIGHSHSNYELFKPEGVAKKEGKNKYVTGEEITNVVNLKNSTFNNQN 613
Db* 421 HHADPIDEHKPVGIGHSHSNYELFKPEGVAKKEGKNKYVTGEEITNVVNLKNSTFNNQN 480
QY 614 FTLANGOKRVVSFFPELEKLGINMLVKLITPDGKVKLEKVGKVGEGVGNIANFELDQ 673
Db 481 FTLANGOKRVVSFFPELEKLGINMLVKLITPDGKVKLEKVGKVGEGVGNIANFELDQ 540
QY 674 PYLPQGTFRYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPEHAGDTYLRVNPQFAV 733
Db 541 PYLPQGTFRYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPEHAGDTYLRVNPQFAV 600
QY 734 PKGTALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTFMANAYLDN 793
Db 601 PKGTALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTFMANAYLDN 660
QY 794 QSTYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 853
Db 661 QSTYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGN 720
QY 854 STSNSTLEEVPTVDVQVKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFT 913
Db 721 STSNSTLEEVPTVDVQVKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFT 780
QY 914 GEAPOGNGENKPSSENGKYSTGVENQPTENKPADSLPEAPNEKPKVKNPENSTDNGLNPEG 973
Db 781 GEAPOGNGENKPSSENGKYSTGVENQPTENKPADSLPEAPNEKPKVKNPENSTDNGLNPEG 840
QY 974 NVGSDPMLDPALEAPAVDPVQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKN 1033
Db 841 NVGSDPMLDPALEAPAVDPVQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKN 900

QY 1034 LSDFIA 1039
Db 901 LSDFIA 906
```

```
RESULT 45
US-09-884-465A-256
; Sequence 256, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 053190-0044
; CURRENT APPLICATION NUMBER: US/09/884.465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 256
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
; OTHER INFORMATION: Unknown Sequence
US-09-884-465A-256

Query Match          46.8%; Score 484; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 556 ADPTDEHKPVGIGHSHSNYELFKPEGVAKKEGKNKYVTGEEITNVVNLKNSTFNNQNT 615
Db 86 ADPTDEHKPVGIGHSHSNYELFKPEGVAKKEGKNKYVTGEEITNVVNLKNSTFNNQNT 145
QY 616 LANGOKRVVSFFPELEKLGINMLVKLITPDGKVKLEKVGKVGEGVGNIANFELDQPY 675
Db 146 LANGOKRVVSFFPELEKLGINMLVKLITPDGKVKLEKVGKVGEGVGNIANFELDQPY 205
QY 676 LPQGTFRYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPEHAGDTYLRVNPQFAVK 735
Db 206 LPQGTFRYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPEHAGDTYLRVNPQFAVK 265
QY 736 GTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTFMANAYLDNOS 795
Db 266 GTDALVRVDFEFGHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTFMANAYLDNOS 325
QY 796 TYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNST 855
Db 326 TYIIVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNST 385
QY 856 SNSTLEEVPTVDVQVKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGE 915
Db 386 SNSTLEEVPTVDVQVKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGE 445
QY 916 APOGNGENKPSSENGKYSTGVENQPTENKPADSLPEAPNEKPKVKNPENSTDNGLNPEGNV 975
Db 446 APOGNGENKPSSENGKYSTGVENQPTENKPADSLPEAPNEKPKVKNPENSTDNGLNPEGNV 505
QY 976 GSDPMLDPALEAPAVDPVQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLS 1035
Db 506 GSDPMLDPALEAPAVDPVQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLS 565

QY 1036 DFIA 1039
Db 566 DFIA 569
```

Search completed: May 13, 2003, 14:02:58
Job time : 54 secs

GenCore version 5.1.1.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein. - protein search, using sw model

Run On: May 13, 2003, 13:45:42 ; Search time 22.5622 Seconds
(without alignments)
3118.326 Million cell updates/sec

Title: US-09-471-255-10
Perfect score: 2746
Sequence: 1 MKDLKKIEBKAGIMKQYG.....IELRLPSGEVKKNLSDFIA 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2746	100.0	528	21	AA12719
2	2746	100.0	528	23	AAU84024
3	2746	100.0	568	21	AA12731
4	2746	100.0	568	23	AAU83999
5	2746	100.0	626	23	AAU84018
6	2746	100.0	632	23	AAU83998
7	2746	100.0	632	23	AAU84002
8	2746	100.0	632	23	AAU84003
9	2746	100.0	632	23	AAU84013
10	2746	100.0	632	23	AAU84014

11	2746	100.0	644	23	AAU83996	Truncated variant
12	2746	100.0	807	23	AAU83997	Truncated variant
13	2746	100.0	895	23	AAU84070	S. pneumoniae deri
14	2746	100.0	896	23	AAU84080	S. pneumoniae deri
15	2746	100.0	901	23	AAU84068	S. pneumoniae deri
16	2746	100.0	901	23	AAU84072	S. pneumoniae deri
17	2746	100.0	901	23	AAU84074	S. pneumoniae deri
18	2746	100.0	902	23	AAU84076	S. pneumoniae deri
19	2746	100.0	902	23	AAU84078	S. pneumoniae deri
20	2746	100.0	913	23	AAU84059	S. pneumoniae deri
21	2746	100.0	999	23	AAU84051	S. pneumoniae deri
22	2746	100.0	999	23	AAU84052	S. pneumoniae deri
23	2746	100.0	1019	21	AA12722	Streptococcus pneu
24	2746	100.0	1019	21	AA12751	Streptococcus pneu
25	2746	100.0	1019	23	AAU84021	Truncated variant
26	2746	100.0	1039	21	AA12715	Streptococcus pneu
27	2746	100.0	1039	23	AAU75932	Streptococcus pneu
28	2746	100.0	1057	21	AA12725	Streptococcus pneu
29	2746	100.0	1058	23	AAU84097	S. pneumoniae deri
30	2746	100.0	1126	23	AAU84058	S. pneumoniae deri
31	2746	100.0	1139	23	AAU84055	S. pneumoniae deri
32	2746	100.0	1152	23	AAU84054	S. pneumoniae deri
33	2746	100.0	1238	23	AAU84056	S. pneumoniae deri
34	2746	100.0	1365	23	AAU84057	S. pneumoniae deri
35	2746	100.0	1378	23	AAU84053	S. pneumoniae deri
36	2740	99.8	1019	21	AA12749	Streptococcus pneu
37	2737	99.7	632	23	AAU84004	Truncated variant
38	2737	99.7	632	23	AAU84006	Truncated variant
39	2737	99.7	632	23	AAU84009	Truncated variant
40	2737	99.7	632	23	AAU84012	Truncated variant
41	2737	99.7	901	23	AAU84066	S. pneumoniae deri
42	2737	99.7	901	23	AAU84067	S. pneumoniae deri
43	2737	99.7	907	23	AAU84060	S. pneumoniae deri
44	2737	99.7	907	23	AAU84061	S. pneumoniae deri
45	2737	99.7	907	23	AAU84064	S. pneumoniae deri

ALIGNMENTS

RESULT 1
AAB12719
ID AAB12719 standard; Protein; 528 AA.
XX AAB12719;
XX

21-NOV-2000 (first entry)

Streptococcus pneumoniae BVH-3B protein antigen SEQ ID NO:10.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

WO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

N-PSDB; AAA65734.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
otitis media; bacteraemia and/or pneumonia

```
XX PS Claim 18; Fig 10; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3 protein antigen.
XX
SQ Sequence 528 AA;
Query Match 100.0%; Score 2746; DB 21; Length 528;
Best Local Similarity 100.0%; Pred. No. 2.4e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKDLKKIEEKTAGIMKQYGVKRESIVVVKENKAIYPHGDHHAADPIDEHKPVGICHSH 60
DB 1 MKDLKKIEEKTAGIMKQYGVKRESIVVVKENKAIYPHGDHHAADPIDEHKPVGICHSH 60
QY 61 SNYELFKPEGVAKGKGNKVTGELTNVNLKSTNFTLANGQKRVSFSPPEL 120
DB 61 SNYELFKPEGVAKGKGNKVTGELTNVNLKSTNFTLANGQKRVSFSPPEL 120
QY 121 EKKLGIMLVKLIITPDGKVLKSVKGVGEGVGNIAFELDQYLPFGTFRYTIASKDYP 180
DB 121 EKKLGIMLVKLIITPDGKVLKSVKGVGEGVGNIAFELDQYLPFGTFRYTIASKDYP 180
QY 181 EYSYDGTFTVPSLAYKMASQIFYPFHAGDTYLRVNPQFAVPKGTDLVRVDFEHGNA 240
DB 181 EYSYDGTFTVPSLAYKMASQIFYPFHAGDTYLRVNPQFAVPKGTDLVRVDFEHGNA 240
QY 241 YLENNYKVGKELIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTD 300
DB 241 YLENNYKVGKELIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTD 300
QY 301 KPSILPQFRNKQAQENSKLDEKVEEPKTSKVEKEKLSGTGNSNSTLEEVPTVDPVQVE 360
DB 301 KPSILPQFRNKQAQENSKLDEKVEEPKTSKVEKEKLSGTGNSNSTLEEVPTVDPVQVE 360
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVYIKKNMADFTGEAPQNGENKPSGENKV 420
DB 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVYIKKNMADFTGEAPQNGENKPSGENKV 420
QY 421 STGTVENQPTENKPADSLPEAPNEKVPKPSNSTDNGMLNPEGNVGSDDPMLDPALEAPAV 480
DB 421 STGTVENQPTENKPADSLPEAPNEKVPKPSNSTDNGMLNPEGNVGSDDPMLDPALEAPAV 480
QY 481 DPVQEKLEKFTASYGLDVSIFNMDGTIELRPSGEVYIKKNLSDFIA 528
DB 481 DPVQEKLEKFTASYGLDVSIFNMDGTIELRPSGEVYIKKNLSDFIA 528
RESULT 2
AAU84024
ID AAU84024 standard; Peptide: 528 AA.
XX
AC AAU84024;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-3, BVH-3b.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; muten.
XX
OS Streptococcus pneumoniae.
XX
PS Synthetic.
XX
PN WO200198334-A2.
```

```
XX PD 27-DEC-2001.
XX
XX 19-JUN-2001: 2001WO-CA00908.
XX
XX 20-JUN-2000: 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia.
XX
XX Example 1; Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
XX Staphylococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a truncate
XX of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
XX described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX
XX SQ Sequence 528 AA;
Query Match 100.0%; Score 2746; DB 23; Length 528;
Best Local Similarity 100.0%; Pred. No. 2.4e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKDLKKIEEKTAGIMKQYGVKRESIVVVKENKAIYPHGDHHAADPIDEHKPVGICHSH 60
DB 1 MKDLKKIEEKTAGIMKQYGVKRESIVVVKENKAIYPHGDHHAADPIDEHKPVGICHSH 60
QY 61 SNYELFKPEGVAKGKGNKVTGELTNVNLKSTNFTLANGQKRVSFSPPEL 120
DB 61 SNYELFKPEGVAKGKGNKVTGELTNVNLKSTNFTLANGQKRVSFSPPEL 120
QY 121 EKKLGIMLVKLIITPDGKVLKSVKGVGEGVGNIAFELDQYLPFGTFRYTIASKDYP 180
DB 121 EKKLGIMLVKLIITPDGKVLKSVKGVGEGVGNIAFELDQYLPFGTFRYTIASKDYP 180
QY 181 EYSYDGTFTVPSLAYKMASQIFYPFHAGDTYLRVNPQFAVPKGTDLVRVDFEHGNA 240
DB 181 EYSYDGTFTVPSLAYKMASQIFYPFHAGDTYLRVNPQFAVPKGTDLVRVDFEHGNA 240
QY 241 YLENNYKVGKELIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTD 300
DB 241 YLENNYKVGKELIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTD 300
QY 301 KPSILPQFRNKQAQENSKLDEKVEEPKTSKVEKEKLSGTGNSNSTLEEVPTVDPVQVE 360
DB 301 KPSILPQFRNKQAQENSKLDEKVEEPKTSKVEKEKLSGTGNSNSTLEEVPTVDPVQVE 360
```

```
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVYKKNMADFTGEAPQNGENKPSSENGKV 420
|||||
Db 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVYKKNMADFTGEAPQNGENKPSSENGKV 420
|||||
QY 421 STGTVENOPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 480
|||||
Db 421 STGTVENOPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 480
|||||
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 528
|||||
Db 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 528
|||||

RESULT 3
AAB12731
ID AAB12731 standard; Protein: 568 AA.
XX
AC AAB12731;
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae NEW1 protein antigen SEQ ID NO:64.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur Br, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
XX
PS Claim 18; Fig 29; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae NEW1 protein antigen.
XX
SQ Sequence 568 AA:
Query Match 100.0%; Score 2746; DB 21; Length 568;
Best Local Similarity 100.0%; Pred. No. 2.6e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKEETAGIMKQYGVKRESIVVKNKEKNAIYPHGDDHHPADPIDEKPVGIGHSH 60
|||||
Db 41 MKDLKKEETAGIMKQYGVKRESIVVKNKEKNAIYPHGDDHHPADPIDEKPVGIGHSH 100
|||||
QY 61 SNYELFKPEEGVAKGKNKVTGELTNVNLKNSFNQNFNLANGQKRVSFPPPEL 120
|||||
Db 101 SNYELFKPEEGVAKGKNKVTGELTNVNLKNSFNQNFNLANGQKRVSFPPPEL 160
|||||
QY 121 EKKLGINMLVKLITPDGKVLKESKVGEGVGNIANFELDQPYLPQGTFKYTTASKDYP 180
|||||
```

```
Db 161 EKKLGINMLVKLITPDGKVLKESKVGEGVGNIANFELDQPYLPQGTFKYTTASKDYP 220
|||||
QY 181 EYSDGTFTVPTSLAYKVASQTIYYPFHAGDTYLRVNPQFAVPKGTDALYRVFDEFHGNA 240
|||||
Db 221 EYSDGTFTVPTSLAYKVASQTIYYPFHAGDTYLRVNPQFAVPKGTDALYRVFDEFHGNA 280
|||||
QY 241 YLENNYKVGELKPIPKLNOGTTTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKNOTD 300
|||||
Db 281 YLENNYKVGELKPIPKLNOGTTTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKNOTD 340
|||||
QY 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKSETGNSTNSNLTSEVPVDPVQE 360
|||||
Db 341 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKSETGNSTNSNLTSEVPVDPVQE 400
|||||
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVYKKNMADFTGEAPQNGENKPSSENGKV 420
|||||
Db 401 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVYKKNMADFTGEAPQNGENKPSSENGKV 460
|||||
QY 421 STGTVENOPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 480
|||||
Db 461 STGTVENOPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 520
|||||
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 528
|||||
Db 521 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVYKKNLSDFIA 568
|||||

RESULT 4
AAU83999
ID AAU83999 standard; Peptide: 568 AA.
XX
AC AAU83999;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-3, NEW1.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; Streptococcal bacterial infection; mutant; mutain.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
PN WO200198334-A2.
XX
PD 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
PR 20-JUN-2000; 2000US-212683P.
XX
PA (SHIR-) SHIRE BIOCHEM INC.
XX
PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
DR WPI; 2002-122272/16.
XX
PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX
PS Claim 1; Page 1; 113pp; English.
XX
CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
```

Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.

Note: This sequence does not appear in the specification but has been created according to information given in the invention.

XX Sequence 568 AA;

Query Match 100.0%; Score 2746; DB 23; Length 568;
Best Local Similarity 100.0%; Pred. No. 2.6e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLDKTEEKIAGIMKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGSH 60
DB 41 MKDLDKTEEKIAGIMKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGSH 100
QY 61 SNYELFKPEGVAKKGNKYVTGEELTNVNLKSTFNQNTLANGOKRVSFSPPEL 120
DB 101 SNYELFKPEGVAKKGNKYVTGEELTNVNLKSTFNQNTLANGOKRVSFSPPEL 160
QY 121 EKKLGIMLVKLIIPDCKVLEKSGVFGVGVGNIANFELDQYLPQGFXYTIASKDYP 180
DB 161 EKKLGIMLVKLIIPDCKVLEKSGVFGVGVGNIANFELDQYLPQGFXYTIASKDYP 220
QY 181 EYSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAPVKGTDLVRVDFEHGNA 240
DB 221 EYSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAPVKGTDLVRVDFEHGNA 280
QY 241 YLENNYKVEIKLPIPLNQGTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENQTD 300
DB 281 YLENNYKVEIKLPIPLNQGTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENQTD 340
QY 301 KPSILPQFKRNKAGENSKLDEKVEEPTSEKVEKLSGTGNTSTNLTVEEPTVDPVQE 360
DB 341 KPSILPQFKRNKAGENSKLDEKVEEPTSEKVEKLSGTGNTSTNLTVEEPTVDPVQE 400
QY 361 KVAFKESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQGNENKPSNGKV 420
DB 401 KVAFKESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQGNENKPSNGKV 460
QY 421 STGIVENQPTENKPADSLPEAPNEKPKVKNPNSDNGMLNPEGNVGSDDMLDPALEAPAV 480
DB 461 STGIVENQPTENKPADSLPEAPNEKPKVKNPNSDNGMLNPEGNVGSDDMLDPALEAPAV 520
QY 481 DPVQEKLEKTASYGLDGVIFNMDGTIELRLPSGEVIAKKNLSDFIA 528
DB 521 DPVQEKLEKTASYGLDGVIFNMDGTIELRLPSGEVIAKKNLSDFIA 568

RESULT 5

AAU84018

ID AAU84018 standard; Peptide; 626 AA.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX WO200198334-A2.
XX 27-DEC-2001.
XX 19-JUN-2001: 2001WC-CA00908.
XX 20-JUN-2000: 2000US-212683P.
XX (SHIR-) SHIRE BIOCHEM INC.
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI: 2002-122272/16.
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
XX Claim 1: Page 1; 113pp; English.
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or
XX Staphylococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumoniae nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a truncate
XX of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
XX described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.

XX Sequence 626 AA;

Query Match 100.0%; Score 2746; DB 23; Length 626;
Best Local Similarity 100.0%; Pred. No. 3e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLDKTEEKIAGIMKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGSH 60
DB 99 MKDLDKTEEKIAGIMKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGSH 158
QY 61 SNYELFKPEGVAKKGNKYVTGEELTNVNLKSTFNQNTLANGOKRVSFSPPEL 120
DB 159 SNYELFKPEGVAKKGNKYVTGEELTNVNLKSTFNQNTLANGOKRVSFSPPEL 216
QY 121 EKKLGIMLVKLIIPDCKVLEKSGVFGVGVGNIANFELDQYLPQGFXYTIASKDYP 180
DB 219 EKKLGIMLVKLIIPDCKVLEKSGVFGVGVGNIANFELDQYLPQGFXYTIASKDYP 278
QY 181 EYSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAPVKGTDLVRVDFEHGNA 240
DB 279 EYSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAPVKGTDLVRVDFEHGNA 338
QY 241 YLENNYKVEIKLPIPLNQGTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENQTD 300
DB 339 YLENNYKVEIKLPIPLNQGTRTAGNKIPVTFMANAYLDNQSTYIVVEPILKENQTD 398
QY 301 KPSILPQFKRNKAGENSKLDEKVEEPTSEKVEKLSGTGNTSTNLTVEEPTVDPVQE 360

Db 399 KPSILPQKRRKNAQENSKLDEKVEPKTSEKVEKELSETGNSTNSTLEEYPTVDPVQE 458
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNWADFTGEAPQNGENKPSGENKV 420
Db 459 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNWADFTGEAPQNGENKPSGENKV 518
QY 421 STGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 480
Db 519 STGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 578
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 528
Db 579 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 626

RESULT 6

AAU83998

ID AAU83998 standard; Peptide: 632 AA.

XX AC AAU83998;
XX DT
XX XX

08-MAY-2002 (first entry)

Truncated variant of *S. pneumoniae* BVH-11, NEW40.XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.
XXOS Streptococcus pneumoniae.
OS Synthetic.

XX WO200198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PI
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.

XX

PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia
XX

PS Example 1; Page 1; 113pp; English.

CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for *S. pneumoniae* infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating *S. pneumoniae* nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

XX SO Sequence 632 AA;
XXQuery Match 100.0%; Score 2746; DB 23; Length 632;
Best Local Similarity 100.0%; Pred. No. 3e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLDDKIEEKIAGIMKQYGVKRESIVVNVKKNIAIYPHGDHHDHAFIDEHKPVGIGHSH 60
Db 105 MKDLDDKIEEKIAGIMKQYGVKRESIVVNVKKNIAIYPHGDHHDHAFIDEHKPVGIGHSH 164
QY 61 SNYLEKPEEGVAKKQYVVTGELTNVNLKNSFNQNTFLANGQKRVSFSPPEL 120
Db 165 SNYLEKPEEGVAKKQYVVTGELTNVNLKNSFNQNTFLANGQKRVSFSPPEL 224
QY 121 EKKIGINMLVKLIIPDGKVKLEKVGKVGEGVGNIAFELDDQVLPQGTFKYTIASKDYP 180
Db 225 EKKIGINMLVKLIIPDGKVKLEKVGKVGEGVGNIAFELDDQVLPQGTFKYTIASKDYP 284
QY 181 EVSYDGTFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAVPGKIDALVRVDFEFGNA 240
Db 285 EVSYDGTFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAVPGKIDALVRVDFEFGNA 344
QY 241 YLENNYKVGGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKNOTD 300
Db 345 YLENNYKVGGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKNOTD 404
QY 301 KPSILPQKRRKNAQENSKLDEKVEPKTSEKVEKELSETGNSTNSTLEEYPTVDPVQE 360
Db 405 KPSILPQKRRKNAQENSKLDEKVEPKTSEKVEKELSETGNSTNSTLEEYPTVDPVQE 464
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNWADFTGEAPQNGENKPSGENKV 420
Db 465 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNWADFTGEAPQNGENKPSGENKV 524
QY 421 STGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 480
Db 525 STGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 584
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 528
Db 585 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 632

RESULT 7

AAU84002

ID AAU84002 standard; Peptide: 632 AA.

XX AC AAU84002;
XX DT

08-MAY-2002 (first entry)

Truncated variant of *S. pneumoniae* BVH-3, NEW49.XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.
XXOS Streptococcus pneumoniae.
OS Synthetic.XX WO200198334-A2.
XX PD

XX 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PIXX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX

XX WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

PT epitope-bearing polypeptides, useful as vaccine components for treating

PT or preventing streptococcal infections such as otitis media,

XX meningitis, and bacteraemia

PS Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or

CC Staphylococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate

CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

XX Sequence 632 AA;

Query Match 100.0%; Score 2746; DB 23; Length 632;

Best Local Similarity 100.0%; Pred. No. 3e-174;

Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIKQYGVKRESIVVKNKNAIIPHGHHADPIDERKPVGIGHSH 60

DB 105 MKDLKKIEEKIAGIKQYGVKRESIVVKNKNAIIPHGHHADPIDERKPVGIGHSH 164

QY 61 SNYELFPEGVAKKGNKYVTGEELTNVNNLKNSTFNQNTLANGKRVSFPPPEL 120

DB 165 SNYELFPEGVAKKGNKYVTGEELTNVNNLKNSTFNQNTLANGKRVSFPPPEL 224

QY 121 EKLGINMLVKLIIPDGKVLKRVSGVFGVGVGNIAFELDPQLPGQTFKVTIASKDYP 180

DB 225 EKLGINMLVKLIIPDGKVLKRVSGVFGVGVGNIAFELDPQLPGQTFKVTIASKDYP 284

QY 181 EYSYDGNFTVTSIAYKMASOTIFYPFHAGDTVLRVNPQFAPVKCTDALRVDFEFGNA 240

DB 285 EYSYDGNFTVTSIAYKMASOTIFYPFHAGDTVLRVNPQFAPVKCTDALRVDFEFGNA 344

QY 241 YLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVPILEKENQTD 300

DB 345 YLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVPILEKENQTD 404

QY 301 KPSILPQFKRNKAENSKLDEKVEEPTSEKVEKLSGTGNTSTNLTLEVPVDPVQVE 360

DB 405 KPSILPQFKRNKAENSKLDEKVEEPTSEKVEKLSGTGNTSTNLTLEVPVDPVQVE 464

QY 361 KVAFASVGMKLENLVNFDGTLIELYLPSPGEVVKKNMADFTCEAPQNGENKPSNGKV 420

DB 465 KVAFASVGMKLENLVNFDGTLIELYLPSPGEVVKKNMADFTCEAPQNGENKPSNGKV 524

QY 421 STGVENQPTENKPADSLPEAPKPKVPKPNSTDNGLNPEGVNGSDPMDLPALEAPAV 480

DB 525 STGVENQPTENKPADSLPEAPKPKVPKPNSTDNGLNPEGVNGSDPMDLPALEAPAV 584

QY 481 DPVQEKLEKFTASYGLDVSIFNMDSITIELRLPSGEVVKKNLSDPIA 528

DB 585 DPVQEKLEKFTASYGLDVSIFNMDSITIELRLPSGEVVKKNLSDPIA 632

RESULT 8

AAU84003

ID AAU84003 standard; Peptide; 632 AA.

XX AC AAU84003;

XX 08-MAY-2002 (first entry)

XX DE Truncated variant of S. pneumoniae BVH-3, NEW50.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KW pneumonia; streptococcal bacterial infection; mutant; mutein.

XX Streptococcus pneumoniae.

OS Synthetic.

XX WO200198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA00908.

XX 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/16.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

PT epitope-bearing polypeptides, useful as vaccine components for treating

PT or preventing streptococcal infections such as otitis media,

PT meningitis, and bacteraemia

XX Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or

CC Staphylococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for S. pneumoniae infection. (III) is useful for designing DNA

CC probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be

CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate

CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

CC described in the method of the invention.

CC Note: This sequence does not appear in the specification but has

CC been created according to information given in the invention.

XX Sequence 632 AA;

Query Match 100.0%; Score 2746; DB 23; Length 632;

Best Local Similarity 100.0%; Pred. No. 3e-174;

Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIKQYGVKRESIVVKNKNAIIPHGHHADPIDERKPVGIGHSH 60

DB 105 MKDLKKIEEKIAGIKQYGVKRESIVVKNKNAIIPHGHHADPIDERKPVGIGHSH 164

QY 61 SNYELFPEGVAKKGNKYVTGEELTNVNNLKNSTFNQNTLANGKRVSFPPPEL 120

|||||
Db 165 SNYELFKPEEGVAKKQVYTGELINNVNLLNSTFNQNF LANGQRVSFSPPEL 244
QY 121 EKKLGINMLVKLITPDGKVLKSKVFGGCVGNIANFELDPYLPQGTQTKYTIASKDYP 180
Db 225 EKKLGINMLVKLITPDGKVLKSKVFGGCVGNIANFELDPYLPQGTQTKYTIASKDYP 284
QY 181 EVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 240
Db 285 EVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 344
QY 241 YLENNYKVGTEIKLPIKLNQGTTFAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTD 300
Db 345 YLENNYKVGTEIKLPIKLNQGTTFAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTD 404
QY 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTVDVQJE 360
Db 405 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTVDVQJE 464
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPQGTQTKYTIASKDYP 420
Db 465 KVAKFAESYGMKLENVLFNMDGTIELYLPQGTQTKYTIASKDYP 524
QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNGVSDPMLDPALEAPAV 480
Db 525 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNGVSDPMLDPALEAPAV 584
QY 481 DPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 528
Db 585 DPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 632

RESULT 9
AAU84013 ID AAU84013 standard; Peptide: 632 AA.
XX AC AAU84013;
XX DT 08-MAY-2002 (first entry)
XX DE Truncated variant of S. pneumoniae BVH-3, NEW63.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX OS pneumonia: streptococcal bacterial infection; mutant; mutein.
XX OS Streptococcus pneumoniae.
XX OS Synthetic.
XX PN WO200198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia

Claim 1: Page -: 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of

meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.
Note: This sequence does not appear in the specification but has been created according to information given in the invention.
XX XX
SQ Sequence 632 AA;

Query Match 100.0%; Score 2746; DB 23; Length 632;
Best Local Similarity 100.0%; Pred. No. 3e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEKTAGINKQYGVKRESIVVANKENAIYPHGDHHPIDEHKPVGIGHSH 60
Db 105 MKDLKKIEKTAGINKQYGVKRESIVVANKENAIYPHGDHHPIDEHKPVGIGHSH 164
QY 61 SNYELFKPEEGVAKKQVYTGELINNVNLLKSTFNQNF LANGQRVSFSPPEL 120
Db 165 SNYELFKPEEGVAKKQVYTGELINNVNLLKSTFNQNF LANGQRVSFSPPEL 224
QY 121 EKKLGINMLVKLITPDGKVLKSKVFGGCVGNIANFELDPYLPQGTQTKYTIASKDYP 180
Db 225 EKKLGINMLVKLITPDGKVLKSKVFGGCVGNIANFELDPYLPQGTQTKYTIASKDYP 284
QY 181 EVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 240
Db 285 EVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 344
QY 241 YLENNYKVGTEIKLPIKLNQGTTFAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTD 300
Db 345 YLENNYKVGTEIKLPIKLNQGTTFAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTD 404
QY 301 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTVDVQJE 360
Db 405 KPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTVDVQJE 464
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPQGTQTKYTIASKDYP 420
Db 465 KVAKFAESYGMKLENVLFNMDGTIELYLPQGTQTKYTIASKDYP 524
QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNGVSDPMLDPALEAPAV 480
Db 525 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNGVSDPMLDPALEAPAV 584
QY 481 DPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 528
Db 585 DPVOEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 632

RESULT 10
AAU84014 ID AAU84014 standard; Peptide: 632 AA.
XX AC AAU84014;
XX DT 08-MAY-2002 (first entry)
XX DE Truncated variant of S. pneumoniae BVH-3, NEW64.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KW pneumonia; streptococcal bacterial infection; mutant; mutin.
 XX Streptococcus pneumoniae.
 OS Synthetic.
 PN WO200198334-A2.
 XX 27-DEC-2001.
 PD 19-JUN-2001; 2001WO-CA00908.
 PF 20-JUN-2000; 2000US-212683P.
 PR (SHIR-) SHIRE BIOCHEM INC.
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 DR
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia.
 XX
 PS Claim 1; Page -: 113pp; English.
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX
 SQ Sequence 632 AA;
 Query Match 100.0%; Score 2746; DB 23; Length 632;
 Best Local Similarity 100.0%; Pred. No. 3e-174;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKDLKKIEKTAGIMKQYKRSIVVKNCKNATYIPGHGHHADPIDEHKPVGIGHSH 60
 DB 105 MKDLKKIEKTAGIMKQYKRSIVVKNCKNATYIPGHGHHADPIDEHKPVGIGHSH 164
 QY 61 SNYELFKPEEGVAKKGNKVTGESLTNVALLKNSTNNQFTLANGQKRVSFPEL 120
 DB 165 SNYELFKPEEGVAKKGNKVTGESLTNVALLKNSTNNQFTLANGQKRVSFPEL 224
 QY 121 EKKLGIMLVKLIITPDGKVLKRSKVGEGVGNIANFELDOPVLPQGTFFYTKASRDYP 180
 DB 225 EKKLGIMLVKLIITPDGKVLKRSKVGEGVGNIANFELDOPVLPQGTFFYTKASRDYP 284
 QY 181 EYSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAPVPGKTDALVRVDFEFGNA 240
 DB 285 EYSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAPVPGKTDALVRVDFEFGNA 344
 QY 241 YLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLQNSTYIVVEPILKENQTD 300
 DB 345 YLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLQNSTYIVVEPILKENQTD 404
 QY 301 KPSILPQFRKKAQENSKLDEKVEEPKTSEKVEKELSETGNTSNTLEEVPTVDPVQE 360
 DB 405 KPSILPQFRKKAQENSKLDEKVEEPKTSEKVEKELSETGNTSNTLEEVPTVDPVQE 464
 QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQNGENKPSENGKV 420
 DB 465 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQNGENKPSENGKV 524
 QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGVSDPMDPALPEEAPAV 480
 DB 525 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGVSDPMDPALPEEAPAV 584
 QY 481 DPVOEKLEKFTASYGLGLOSVIFNMDGTIELRLPSEGVIKKNLSDFIA 528
 DB 585 DPVOEKLEKFTASYGLGLOSVIFNMDGTIELRLPSEGVIKKNLSDFIA 632

Db 345 YLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLQNSTYIVVEPILKENQTD 404
 QY 301 KPSILPQFRKKAQENSKLDEKVEEPKTSEKVEKELSETGNTSNTLEEVPTVDPVQE 360
 Db 405 KPSILPQFRKKAQENSKLDEKVEEPKTSEKVEKELSETGNTSNTLEEVPTVDPVQE 464
 QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQNGENKPSENGKV 420
 Db 465 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQNGENKPSENGKV 524
 QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGVSDPMDPALPEEAPAV 480
 Db 525 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGVSDPMDPALPEEAPAV 584
 QY 481 DPVOEKLEKFTASYGLGLOSVIFNMDGTIELRLPSEGVIKKNLSDFIA 528
 Db 585 DPVOEKLEKFTASYGLGLOSVIFNMDGTIELRLPSEGVIKKNLSDFIA 632
 RESULT 11
 AAU83996
 ID AAU83996 standard; Peptide; 644 AA.
 XX AAU83996;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Truncated variant of S. pneumoniae BVH-11, NEW21.
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutin.
 XX
 OS Streptococcus pneumoniae.
 CS Synthetic.
 PN WO200198334-A2.
 XX 27-DEC-2001.
 PD 19-JUN-2001; 2001WO-CA00908..
 PF 20-JUN-2000; 2000US-212683P.
 PR (SHIR-) SHIRE BIOCHEM INC.
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 DR
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia.
 XX
 PS Example 1; Page -: 113pp; English.
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for

CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX
 XX
 SQ Sequence 644 AA;

Query Match 100.0%; Score 2746; DB 23; Length 644;
 Best Local Similarity 100.0%; Pred. No. 3.1e-174;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVNKEKNAIIPHGHHADPIDEHKPVGIGHSH 60
 |||||
 Db 117 MKDLKKIEEKIAGIMKQYGVKRESIVVNKEKNAIIPHGHHADPIDEHKPVGIGHSH 176
 QY 61 SNYELFKPEEGVAKKEGKQVYTGELTNVLLKNSIFNNQFTLANGQKRVSFPEPEL 120
 |||||
 Db 177 SNYELFKPEEGVAKKEGKQVYTGELTNVLLKNSIFNNQFTLANGQKRVSFPEPEL 236
 QY 121 EKLGINMLVLIITPDGKVLKESKVGEGVGNITANFELDQPLPGQTFYTTASKDYP 180
 |||||
 Db 237 EKLGINMLVLIITPDGKVLKESKVGEGVGNITANFELDQPLPGQTFYTTASKDYP 296
 QY 181 EYSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 240
 |||||
 Db 297 EYSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 356
 QY 241 YLENNYKVGELIKLIPKLNQGTTRTAGNKIPVTMANAYLDNQSTIYVEVPILKEKQTD 300
 |||||
 Db 357 YLENNYKVGELIKLIPKLNQGTTRTAGNKIPVTMANAYLDNQSTIYVEVPILKEKQTD 416
 QY 301 KPSILPQFRKKAQNSKLDKEVEPKTSEKVEKEKLSSETGNSNSTLEEVVPVDPVQE 360
 |||||
 Db 417 KPSILPQFRKKAQNSKLDKEVEPKTSEKVEKEKLSSETGNSNSTLEEVVPVDPVQE 476
 QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPAGEVTKKNMADFTGEAPQNGENKPSENGKV 420
 |||||
 Db 477 KVAKFAESYGMKLENVLFNMDGTIELYLPAGEVTKKNMADFTGEAPQNGENKPSENGKV 536
 QY 421 STGTVENOPTENKPADSLPEAPNEKFPKVPENSTNGMLNPEGNGVSDPMLDPALEAPAV 480
 |||||
 Db 537 STGTVENOPTENKPADSLPEAPNEKFPKVPENSTNGMLNPEGNGVSDPMLDPALEAPAV 596
 QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGSEVIKKNLSDFFIA 528
 |||||
 Db 597 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGSEVIKKNLSDFFIA 644

RESULT 12

AAU83997
 ID AAU83997 standard; Peptide: 807 AA.

XX
 XX
 AC AAU83997;

XX
 XX
 DT 08-MAY-2002 (first entry)

XX
 DE Truncated variant of S. pneumoniae BVH-11, NEW25.

XX
 KW BVH-3; BVH-11; vaccine; meningitis; Otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutain.

XX
 OS Streptococcus pneumoniae.

XX
 OS Synthetic.

XX
 PN WO200198334-A2.

XX
 PD 27-DEC-2001.

XX
 PF 19-JUN-2001; 2001WO-CA00908.

XX
 PR 20-JUN-2000; 2000US-212683P.

XX

PA (SHIR-) SHIRE BIOCHEM INC.

XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX
 DR WPI; 2002-122272/16.

XX
 XX

PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia

XX
 PS Example 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides.
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX
 SQ Sequence 807 AA;

Query Match 100.0%; Score 2746; DB 23; Length 807;
 Best Local Similarity 100.0%; Pred. No. 4.1e-174;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVNKEKNAIIPHGHHADPIDEHKPVGIGHSH 60
 |||||
 Db 280 MKDLKKIEEKIAGIMKQYGVKRESIVVNKEKNAIIPHGHHADPIDEHKPVGIGHSH 339
 QY 61 SNYELFKPEEGVAKKEGKQVYTGELTNVLLKNSIFNNQFTLANGQKRVSFPEPEL 120
 |||||
 Db 340 SNYELFKPEEGVAKKEGKQVYTGELTNVLLKNSIFNNQFTLANGQKRVSFPEPEL 399
 QY 121 EKLGINMLVLIITPDGKVLKESKVGEGVGNITANFELDQPLPGQTFYTTASKDYP 180
 |||||
 Db 400 EKLGINMLVLIITPDGKVLKESKVGEGVGNITANFELDQPLPGQTFYTTASKDYP 459
 QY 181 EYSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 240
 |||||
 Db 460 EYSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 519
 QY 241 YLENNYKVGELIKLIPKLNQGTTRTAGNKIPVTMANAYLDNQSTIYVEVPILKEKQTD 300
 |||||
 Db 520 YLENNYKVGELIKLIPKLNQGTTRTAGNKIPVTMANAYLDNQSTIYVEVPILKEKQTD 579
 QY 301 KPSILPQFRKKAQNSKLDKEVEPKTSEKVEKEKLSSETGNSNSTLEEVVPVDPVQE 360
 |||||
 Db 580 KPSILPQFRKKAQNSKLDKEVEPKTSEKVEKEKLSSETGNSNSTLEEVVPVDPVQE 639
 QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPAGEVTKKNMADFTGEAPQNGENKPSENGKV 420
 |||||
 Db 640 KVAKFAESYGMKLENVLFNMDGTIELYLPAGEVTKKNMADFTGEAPQNGENKPSENGKV 699
 QY 421 STGTVENOPTENKPADSLPEAPNEKFPKVPENSTNGMLNPEGNGVSDPMLDPALEAPAV 480
 |||||
 Db 700 STGTVENOPTENKPADSLPEAPNEKFPKVPENSTNGMLNPEGNGVSDPMLDPALEAPAV 759

QY	481	DPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVKKNLSDFIA	528
DB	760	DPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVKKNLSDFIA	807
RESULT 13			
AAU84070	AAU84070 standard; Peptide: 895 AA.		
XX	AC	AAU84070;	
XX	DT	08-MAY-2002 (first entry)	
XX	DE	S. pneumoniae derived chimeric peptide, VP112.	
XX	KW	BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;	
XX	KW	pneumonia; streptococcal bacterial infection; mutant; muteln;	
XX	KW	BVH-11-2.	
XX	OS	Streptococcus pneumoniae.	
XX	OS	Synthetic.	
XX	PN	WO200198334-A2.	
XX	PD	27-DEC-2001.	
XX	PF	19-JUN-2001; 2001WO-CA00908.	
XX	PR	20-JUN-2000; 2000US-212683P.	
XX	PA	(SHIR-) SHIRE BIOCHEM INC.	
XX	PI	Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;	
XX	DR	WPI; 2002-122272/16.	
XX	PS	Claim 1; Page -; 113pp; English.	
XX	CC	The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacardia or Streptococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.	
XX	CC	Note: This sequence does not appear in the specification but has been created according to information given in the invention.	
XX	CC	Sequence	895 AA;
XX	CC	Query Match	100.0%; Score 2746; DB 23; Length 895;
XX	CC	Best Local Similarity	100.0%; Pred. No. 4.7e-174;
XX	CC	Matches 528; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

PS Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.

CC Note: This sequence does not appear in the specification but has been created according to information given in the invention.

XX

SQ Sequence 896 AA:

Query Match 100.0%; Score 2746; DB 23; Length 896;
Best Local Similarity 100.0%; Pred. No. 4.7e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVKNKNAIIPHGDDHADDIDEHKPVGIGHSH 60
DB 369 MKDLKKIEEKIAGIMKQYGVKRESIVVKNKNAIIPHGDDHADDIDEHKPVGIGHSH 428

QY 61 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKNSFNQNTLANGOKRVSFPPEL 120
DB 429 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKNSFNQNTLANGOKRVSFPPEL 488

QY 121 EKKGINMLVKLITPDGKVLKESKVGKVGEGVGNIANFELDQYLPFGOTFKYTIASKDYP 180
DB 489 EKKGINMLVKLITPDGKVLKESKVGKVGEGVGNIANFELDQYLPFGOTFKYTIASKDYP 548

QY 181 EYSYDGTFTVPTSLAYKASQTIFFPHAGDTYLRVNPQFAVPKGTDALRVDFEFHNA 240
DB 549 EYSYDGTFTVPTSLAYKASQTIFFPHAGDTYLRVNPQFAVPKGTDALRVDFEFHNA 608

QY 241 YLENNKYVGEIKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIIVEPILKRNQTD 300
DB 609 YLENNKYVGEIKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIIVEPILKRNQTD 668

QY 301 KPSILPQKRNKAQENSKLDERVEPKSEKVEKEKLSGTGNSNSTSLIEEYPTVDPVQE 360
DB 669 KPSILPQKRNKAQENSKLDERVEPKSEKVEKEKLSGTGNSNSTSLIEEYPTVDPVQE 728

QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIRKKNMADFTGEAPQNGENKPSGK 420
DB 729 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIRKKNMADFTGEAPQNGENKPSGK 788

QY 421 STGTVENQTEKNKPADSLPEANEXPKPENSTDNGLNPNENGVSDPMLDPALEAPAV 480
DB 789 STGTVENQTEKNKPADSLPEANEXPKPENSTDNGLNPNENGVSDPMLDPALEAPAV 848

QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 528
DB 849 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIRKKNLSDFIA 896

RESULT 15
AAU84068
ID AAU84068 standard; Peptide; 901 AA.
XX
AC AAU84068;

XX 08-MAY-2002 (first entry)

XX S. pneumoniae derived chimeric peptide, VP110.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KW pneumonia; streptococcal bacterial infection; mutant; mutein;

KW BVH-11-2.

XX Streptococcus pneumoniae.

OS Synthetic.

XX WO2001198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA00908.

XX 20-JUN-2000; 2000US-212683P.

PR (SHIR-) SHIRE BIOCHEM INC.

PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

PI WPI; 2002-122272/16.

DR

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia

XX

PS Claim 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.

CC Note: This sequence does not appear in the specification but has been created according to information given in the invention.

XX

SQ Sequence 901 AA:

Query Match 100.0%; Score 2746; DB 23; Length 901;
Best Local Similarity 100.0%; Pred. No. 4.7e-174;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEEKIAGIMKQYGVKRESIVVKNKNAIIPHGDDHADDIDEHKPVGIGHSH 60
DB 374 MKDLKKIEEKIAGIMKQYGVKRESIVVKNKNAIIPHGDDHADDIDEHKPVGIGHSH 433

QY 61 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKNSFNQNTLANGOKRVSFPPEL 120
DB 434 SNYELFKPEEGVAKKEGKNVYTGELTNVNLKNSFNQNTLANGOKRVSFPPEL 493

QY 121 EKKGINMLVKLITPDGKVLKESKVGKVGEGVGNIANFELDQYLPFGOTFKYTIASKDYP 180
DB 494 EKKGINMLVKLITPDGKVLKESKVGKVGEGVGNIANFELDQYLPFGOTFKYTIASKDYP 553

```

QY 181 EYSYDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQFAVPKGTDAIVRVDFDEFGNA 240
Db 554 EYSYDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQFAVPKGTDAIVRVDFDEFGNA 613
QY 241 YLENNYKVGGEIKLPIPKLNOGTTTAGNKKIPVTFMANAYLDNQSTYIIVEVPILEKENOTD 300
Db 614 YLENNYKVGGEIKLPIPKLNOGTTTAGNKKIPVTFMANAYLDNQSTYIIVEVPILEKENOTD 673
QY 301 KPSILPQFRKRAQENSKLDEKVEEPKTSSEKVEKLSSETGNSTNSSTLEEVPTVDPVQE 360
Db 674 KPSILPQFRKRAQENSKLDEKVEEPKTSSEKVEKLSSETGNSTNSSTLEEVPTVDPVQE 733
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSSENGKV 420
Db 734 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPSSENGKV 793
QY 421 STGTVENQPTENKPADSLPEAPNEKPVKPEENSTDNMGNPEGNVGSDDPMLDPALEAPAV 480
Db 794 STGTVENQPTENKPADSLPEAPNEKPVKPEENSTDNMGNPEGNVGSDDPMLDPALEAPAV 853
QY 481 DPVQEKLEKFTASYGIGLDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 528
Db 854 DPVQEKLEKFTASYGIGLDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 901

```

Search completed: May 13, 2003, 13:53:34
Job time : 23.5622 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:27 ; Search time 11.9447 Seconds

(without alignments)
4249.498 Million cell updates/sec

Title: US-09-471-255-10

Perfect score: 2746

Sequence: 1 MKDLKKTEEKIAGIMKQYG.....IELRPSGEVKKNSDFIA 528

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2740	99.8	1039	2 D97985	hypothetical prote
2	2734	99.6	1039	2 H95115	conserved hypotet
3	184	6.7	2004	2 F95133	immunoglobulin A1
4	169.5	6.2	1963	2 B98002	IgA-specific metal
5	153	5.6	1873	2 T30944	surface protein pr
6	149.5	5.4	4688	2 F92885	hypothetical prote
7	149	5.4	1849	2 C41859	IgA-specific metal
8	148	5.4	1939	2 T18372	repeat organellar
9	146.5	5.3	791	2 T99924	hypothetical prote
10	146	5.3	1272	2 C90593	hypothetical prote
11	143.5	5.2	1139	1 E64234	cytadherence-acces
12	142	5.2	2485	1 H71621	serine/threonine-s
13	141.5	5.2	1441	2 B66807	hypothetical prote
14	141.5	5.2	1664	2 T18262	S-layer protein -
15	141	5.1	1134	2 A60234	IgA FC receptor pr
16	141	5.1	1164	1 FCSOAG	IgA FC receptor pr
17	139	5.1	750	2 G41361	probable flagellin
18	138.5	5.0	1127	2 T25804	hypothetical prote
19	137.5	5.0	387	2 G97011	alanyl-tRNA synthe
20	137.5	5.0	1526	2 A44406	DNA topoisomerase
21	137	5.0	940	2 A81744	internalin protein
22	136.5	5.0	940	2 AD1374	internalin protein
23	134.5	4.9	2481	2 D90011	FmtB protein [impo
24	134	4.9	380	2 S22415	membrane protein T
25	133	4.8	710	2 S30154	low-temperature-in
26	133	4.8	1020	2 T91018	hypothetical prote
27	132.5	4.8	858	2 S45762	neurofilament trip
28	132	4.8	1231	2 T28661	rib protein - Stre
29	131.5	4.8	1701	2 A26868	major merozoite su

ALIGNMENTS

RESULT 1

D97985

hypothetical protein phrE [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: D97985

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
Y, P.; Sun, P.N.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: D97985

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1039 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:gl5458515; GSPDB:GN00174

C:Genetics:

A:Gene: phrE

Query Match 99.8%; Score 2740; DB 2; Length 1039;

Best Local Similarity 99.8%; Pred. No. 1.8e-150;

Matches 527; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKDLKKTEEKIAGIMKQYGKRESIVVVKENAIYYPHGDDHDPIDEHKPVGIGHSH 60

Db 512 MKDLKKTEEKIAGIMKQYGKRESIVVVKENAIYYPHGDDHDPIDEHKPVGIGHSH 571

Qy 61 SNYELFKPEEGVAKKGNKYVTGELTNVVLKKNSTFNQNTLANGQKRVSFSPPEL 120

Db 572 SNYELFKPEEGVAKKGNKYVTGELTNVVLKKNSTFNQNTLANGQKRVSFSPPEL 631

Qy 121 EKKLGINMLVLIIPDGKVLKSKVFGVGNIANFELDPQVLPQGTFFKTYTASDKYP 180

Db 632 EKKLGINMLVLIIPDGKVLKSKVFGVGNIANFELDPQVLPQGTFFKTYTASDKYP 691

Qy 181 EVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTALVRVDFEHGNA 240

Db 692 EVSDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTALVRVDFEHGNA 751

Qy 241 YLENNYKVGCEIKLPIPKLNOQTTRTAGNKIPVTMANAYLDNQSTYIVEPILKEKQTD 300

Db 752 YLENNYKVGCEIKLPIPKLNOQTTRTAGNKIPVTMANAYLDNQSTYIVEPILKEKQTD 811

Qy 301 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSNSETNSTLEEVPTVDVQVE 360

Db 812 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSNSETNSTLEEVPTVDVQVE 871

Qy 361 KVAFAESYGMKLVNLFNMDGTIELYLPSEGVTKKNAADFTGAPOGNGENKPSSENGKV 420

Db 872 KVAFAESYGMKLVNLFNMDGTIELYLPSEGVTKKNAADFTGAPOGNGENKPSSENGKV 931

glycophorin-bindin
hypothetical prote
hypothetical prote
conserved membrane
collagen adhesin -
ribonucleoside-tri
hypothetical prote
IgA-specific metal
hypothetical prote
histidine Motif-Co
major merozoite su
protein kinase - s
pheromone response
hypothetical prote
fibrinogen-binding
uncharacterized co

30 131 4.8 774 2 A24057
31 130.5 4.8 719 2 AE1131
32 130.5 4.8 853 2 C97985
33 130.5 4.8 896 2 AB1156
34 130.5 4.8 1185 2 A42404
35 130.5 4.8 1750 2 H64403
36 130.5 4.8 5005 2 F82884
37 130 4.7 1702 2 A41859
38 130 4.7 1983 2 G86643
39 129.5 4.7 855 2 D98004
40 129.5 4.7 1086 2 S16752
41 129.5 4.7 1094 2 S49313
42 128.5 4.7 366 2 S56040
43 128.5 4.7 384 2 G86287
44 128.5 4.7 933 2 S41539
45 128.5 4.7 941 2 F97353

```

QY 421 STGTVENOPTENKPADSLPEAPNEKVPKPNSTNGMLNPGNVSDDPMLDPALEAPAV 480
Db 932 STGTVENOPTENKPADSLPEAPNEKVPKPNSTNGMLNPGNVSDDPMLDPALEAPAV 991

QY 481 DPVQKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 528
Db 992 DPVQKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVVKKNLSDLIA 1039

RESULT 2
H95115
conserved hypothetical protein SP1004 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95115
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:g14972477; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
C:Superfamily: Streptococcus sanguis IgA-specific metalloendopeptidase

Query Match 99.6%; Score 2734; DB 2: Length 1039;
Best Local Similarity 99.6%; Pred. No. 3.9e-150;
Matches 526; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKDLKKIEKTAGIMKQYGVKRESIVNKKNAIYPGHGHHADPDIDEHRPVGIGHSH 60
Db 512 MKDLKKIEKTAGIMKQYGVKRESIVNKKNAIYPGHGHHADPDIDEHRPVGIGHSH 571

QY 61 SNYELKPEGVAKKGNKYVTEELTNVNVLLKSTNNQNTFLANGKRVSFSPPEL 120
Db 572 SNYELKPEGVAKKGNKYVTEELTNVNVLLKSTNNQNTFLANGKRVSFSPPEL 631

QY 121 EXKLINMLVKLITPDGKLVKGVKGVGNGNTANFELDPYLPQGTFTYTIASKDYP 180
Db 632 EXKLINMLVKLITPDGKLVKGVKGVGNGNTANFELDPYLPQGTFTYTIASKDYP 691

QY 181 EYSYDGTFTVPTSLAYKMASOTIFYPHAGDTYLRVNPQFAVPKGTDALRVDFDFHNA 240
Db 692 EYSYDGTFTVPTSLAYKMASOTIFYPHAGDTYLRVNPQFAVPKGTDALRVDFDFHNA 751

QY 241 YLENNYKVGITKLPKPLNGQTTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILEKNQTD 300
Db 752 YLENNYKVGITKLPKPLNGQTTTRTAGNKIPVTFMANAYLDNQSTYIVVEPILEKNQTD 811

QY 301 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKELSETGNSSTLSEETVDPVQOE 360
Db 812 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKELSETGNSSTLSEETVDPVQOE 871

QY 361 KVAKFAESYGMKLVNLFNMDGTIELYLPSPGEVVKKNMADFTGEAPQNGENKPSENGKV 420
Db 872 KVAKFAESYGMKLVNLFNMDGTIELYLPSPGEVVKKNMADFTGEAPQNGENKPSENGKV 931

QY 421 STGTVENOPTENKPADSLPEAPNEKVPKPNSTNGMLNPGNVSDDPMLDPALEAPAV 480
Db 932 STGTVENOPTENKPADSLPEAPNEKVPKPNSTNGMLNPGNVSDDPMLDPALEAPAV 991

QY 481 DPVQKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVVKKNLSDFIA 528
Db 992 DPVQKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVVKKNLSDLIA 1039
```

```

RESULT 3
H95133
immunoglobulin A1 proteinase [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 14-Sep-2001
C:Accession: F95133
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95133
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2004 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75263.1; PID:g14972632; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
C:Superfamily: Streptococcus sanguis IgA-specific metalloendopeptidase

Query Match 6.7%; Score 184; DB 2: Length 2004;
Best Local Similarity 23.5%; Pred. No. 0.023;
Matches 130; Conservative 70; Mismatches 162; Indels 192; Gaps 34;

QY 60 HSNYELFKPEGVAKKGNKYVTEELTNVNVLLKSTNNQNTFLANGKRVSFSPPEL 102
Db 279 HKNLET-KKEKISPKETGVTNLNPDQEVLSGOLNKPPELLYREETMETKIDFEEIQEN 337

QY 103 FTLANGKRVSFSPPELKKLGINM-LVKLIT- 144
Db 338 PDLAEGTVRV-----KOEKLGKGVKVEIRFISVKNKEVREIVSTSTTAPSPRIVEKGT 391

QY 145 GK--VFGR--GV-----GNIANFELDQYLPFGOTFKYTIASKDYPEVSDGTFTVP 191
Db 392 KKTQVIREQETGVGHKDVQSGAIVPAI-QPELP-----EAVVSDKCEPEVQP---TLP 442

QY 192 TSLAYKMASOTIFYPHAGDTYLRVNPQ---FAVPKGTDALRVDFDFHGNAYLENNYK 247
Db 443 EAVV-----TDKGET--EVQPESTVTVSDGDEPEQVAPLPEYKGN----- 481

QY 248 VGEIK-LPIPKL-NOGTTTAGNKIPVTFMANAYLDNQSTYIVVEPILEKNQTDKPSI 304
Db 482 LEQVKPEPTVKTKEQPEKT--EEVPV-----KPTETPVNNEGTEGTSI 527

QY 305 LPQKRN---KAQENSKLDEKVEEPTSEK-----VEKEKLESETGNS 343
Db 528 --QEAENPVQAEESTTNSEKV-SPDTSSKNTEGEVSNPNSDSTTSVGEENKPEINDSKNE 584

QY 344 TSNSTLEEVPTVDPVQEKVAKFAESYGMKLVNLFNMDGTIELYLPSPGEVVKKNMADFTG 403
Db 585 NSEKTVEVP-VNP-----NEGIVE-----GTSNQET 610

QY 404 PAPQNGENKPSENGKV---STGTIVENOPTENKIP-ADSLPEAPNEKVPKPNST 454
Db 611 EKPVPQAEETOTNSGKIANENTGEVSNKPSDSKPPVEESNQPEKNGTAKPKNCSNTTSE 670

QY 455 NGMLNPE---GNVGSDDPMLDPALEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIEL 511
Db 671 NSQTEPEFNSNGNSTEDVSTESNTSNGNEIKQENE-----LDPDKKVEPEKTELEL 723

QY 512 RLPSGEVVKKNLS 525
Db 724 R-----NVSD 728

RESULT 4
B98002
IgA-specific metalloendopeptidase (EC 3.4.24.13) [imported] - Streptococcus pneumoniae
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
```


C:Accession: B98002
R: Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
r, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; Mahren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A: Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A: Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A: Reference number: A97872; MUID: 21429245; PMID: 11544234
A: Accession: B98002
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-1963 <KUR>
A: Cross-references: GB:AE007317; PIDN:AAK99846.1; PID:gl5458662; GSPDB:GN00174
C: Geneticks:
C: Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase
C: Keywords: hydrolase, metalloproteinase

Query Match 6.2%; Score 169.5; DB 2; Length 1963;
Best Local Similarity 23.7%; Pred. No. 0.16;
Matches 115; Conservative 60; Mismatches 140; Indels 171; Gaps 30;

QY 60 HSNYELFRPEGVAKKEGKNTYV---GEELTNVYN---LLKNSTFN---NON 102
DB 279 HKNLET-KKEKISPKETGYNTLPQDEVLSGQLNKPPELLYRETIETKIDFQEEIOEN 337
QY 103 FTLANGQRVRSFSPPELEKLGIMN-LVKLIT-----PDGKVLKVS 144
DB 338 PDLAEGTVRV-----KQEGKLGKKEIVRIFSVNKEEVSREIVSTSTTAPSPRIVEKGT 391
QY 145 GK--VEGE---GV-----GNTANFELDOPYLPQGTFKTIASKOYPEVSYDGTETVP 191
DB 392 KKTQVKEQPTGVHEKVDQSGAIVEPAI-OPELP-----EAVSOKGEPEVQ---TLP 442
QY 192 TSLAYKMASQIFYPFHAGDTYLRVNPQ---FAVPKGTDALRVFDEFHGNAYLENNYK 247
DB 443 EAVV-----TDKGET--EVQSPDVTVSDKGEQVAPLPEYKGN-----481
QY 248 VGEIK--LPIPKL-NQGTTRTAGNKIPVTFMANAYLDNQSTIYIEVPILKEKNOTDKPSI 304
DB 482 IEQVAPETPVTEKQGEKPT--BEVVP-----KPTETPVNPNEGTETGSI 527
QY 305 LPQKRN---RAQENSKLDEKVEEPTSEKV-----EKEKLSSETGNT 344
DB 528 --QEAENPVQPAESTTNSEKVSPTDTSSENTGEVSSNPDSSTTSVGSNKPEHNSKNEN 585
QY 345 SNTILEEYPTVDQVQKAKFAESYGMKLENVFNMDGTIELYLPSEGVKKKNMADFTGE 404
DB 586 SEKTVEEVP-VNP-----NEGTVS-----GTSNOETE 611
QY 405 APQNGENKPSNGKV---STGTVENOPTENKP--ADSLPEAPNEKPKVPKPNs---TDN 455
DB 612 KPQVPAETQTSNGKIANENGTGVSNKPSDSKPPVEESNQPEKNGTATKPNsCNTTSEN 671
QY 456 GMLNPE 461
DB 672 GQTEPE 677

RESULT 5
T30944
surface protein precursor - Enterococcus faecalis
C: Species: Enterococcus faecalis
C: Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C: Accession: T30944
R: Shankar, V.; Baghdavan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.
Infect. Immun. 67, 193-200, 1999
A: Title: Infection-derived Enterococcus faecalis strains are enriched in esp, a gene end
A: Reference number: Z20943; MUID: 99081742; PMID: 9864215
A: Accession: T30944
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-1873 <SHA>

A: Cross-references: EMBL:AF034779; NID:g3873186; PID:g3873187; PIDN:AAD09858.1

Query Match 5.6%; Score 153; DB 2; Length 1873;
Best Local Similarity 23.7%; Pred. No. 1.3;
Matches 123; Conservative 58; Mismatches 194; Indels 144; Gaps 28;

QY 2 KDLDKKTI---EKKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHADPIDEHKPVGIG 57
DB 1027 KDLPKDAKYTWKEVD--ISTAGNKKGTVVVT-----YSDGSSDEVE-VD---VTVT 1072
QY 58 HSHSNYELFKPE-EGVAKKEGKNTYVTEELTNVNVLLKNSTFNN-----QNFTLANGQ 109
DB 1073 DNRSADAKYEPTVEGEKVEIGKVDLTNDVNTLPTPOGTTVDVTPGGTIDTNTPGNVE 1132
QY 110 KRYSEFPPPELEKKGIMLVKL---ITPDGKVLKVSQGVFGEV-----GNIAN 157
DB 1133 GVIEVTPD-----GTRKDTVKVPVEVTDNRSADAKYEPTVEGEKVEIGKVDLTNTVNTN 1186
QY 158 FELDQVLPQGTQFYTIASKDYPEVSYDGTFTVPTSILAYKMASOTIFYPHAGDTYLRVN 217
DB 1187 L-----PTLPGT---TVT-----DVTGGTIDTNTPGNEGVIEV-----TV--- 1221
QY 218 PQRAVPKGTDALRVFDEFHGNAYLENNY-----KV---GEIKLP-----IPKLNQGT 262
DB 1222 -----PDGTRDTVKVPVEVTDNRSADAKYEPTVEGEKVEIGKVDLTNDVNTLPTPOGT 1276
QY 263 T---RTAGNKIPVTFMANAYLDNQSTY-----IVEVPILKEKNOTDKPSILPOFKRNKA 313
DB 1277 TVTDVTPGGTIDTNTPGNEGVIEVTPDGTQKDTVKVPVEVTDNRSADAKYTTPWVEGEKV 1336
QY 314 QENSKLDEKVEEPTSEKVEKEKLSSETGNTSNTILEEYPTVDQVQKAKFAESYGMK 373
DB 1337 EIGKVD-----LTDNVNTLPTPOGTTVDVTPG-----GTID 1370
QY 374 ENVLNFMGDTIELYLPSPG---EVIK-----KNMADTGEAPQNGENKPSNGKVS-TGT 424
DB 1371 TNPNGNEGVIEVTPDGTQKDTVKVPVEVTDNRSADAKYEPTVEGE-KVEIGKVDLTND 1429
QY 425 VENQPTENKPADSLPEAPNEKPKVPENSDNGMLNPEGN 463
DB 1430 VTNLPT-----LPGTITVTDVTPGTTIDT---NTPGN 1458

RESULT 6
F82885
hypothetical protein U0482 [imported] - Ureaplasma urealyticum
C: Species: Ureaplasma urealyticum
C: Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C: Accession: F82885
R: Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A: Description: the complete sequence of Ureaplasma urealyticum: Alternate views of a
A: Reference number: A82870
A: Accession: F82885
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-4688 <GLA>
A: Cross-references: GB:AE002145; GB:AF222894; NID:g68999476; PIDN:AAF30894.1; GSPDB:GN
A: Experimental source: serovar 3; biovar 1
C: Geneticks:
A: Gene: U0482
A: Genetic code: SGC3

Query Match 5.4%; Score 149.5; DB 2; Length 4688;
Best Local Similarity 20.6%; Pred. No. 8.1;
Matches 125; Conservative 89; Mismatches 219; Indels 173; Gaps 27;

QY 1 MKDLDKIEKIKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHADPIDEHKPVGIGHSH 60
DB 3859 INDTQQTINVTLSGVNSKYN-GROIKVYKDNVNIY-----ESSLITLQKKG 3905
QY 61 SNVELFKPEEGVAKKEGKNTYVTEELTNVNVLLKNSTFNNQNTLANGQKRVFSFPEL 120

3906 NDYQLL-----LSNLNSNRYREFEKIE--INHISN-TNNFEDEKLNG---VSNFTFITQT 3954
QY 121 EKKL-----GINMLVKLTPDGKVKLEKSGKVGFGVGNANFELDQYLP 166
DB 3955 KNTTQWMDSSATIIVGTGRVAFNEFKIKSD-KILEN-----NQQVAMFAP 3999
QY 167 GQTFK-----YTLASRDYPEVSDGFTVPTSLAYKMASOTIFYPFHAGDTYLRNPQ 219
DB 4000 KETIRDTWLTWYTRPLKDYVSDFKEGTWAHDLNSVNFKEET-----TYKLVKIQ 4050
QY 220 FAVPFGTDLVRVDFRHGNAYLEN-----NY-----KVGEIKL-PIPKLNOQTTRT--- 265
DB 4051 F-VNKPTRAKNNI-NNSENNVLDTNSINSNYEFTTKVGGHKLINITSSNNVNTSOTI 4108
QY 266 -----AGNKIPVTFMANAVLDNQSYIVVEPILKEKNOTDKPSILPOPKRKAQ 314
DB 4109 NFTLSGVKKSWSGKKIKUSKN-----DTSEIHTNEVLIESNKTQYNLLNNLKRNTY 4164
QY 315 ENSKLDKEVEEPTSEKVEKELSTGNSSTLEEVPTVDPVOEKVAKFAESYGMKL- 373
DB 4165 --TLIDVKLIDNNVSDPPKEGNTNSFTITTSAINVLNIEEISNRASTNLKSTLIIN 4222
QY 374 ----ENVLNMDGTIELYLPSEVITKKMAVDTGEAPOGNGENKPSNGKYSTGT---- 424
DB 4223 LNDPDLVRDKDOATIV-----GNKKOAMGFIIVSGNIKYLTAIVLDLN 4267
QY 425 -----VENQPTENKPA-----DSLPEAPNEKPKPENSTD---NGM 457
DB 4268 FNDKYNIVNISFNKPSIAAENIGDKSNIIYNNDISIP-----KLEINNDIIVNGP 4319
QY 458 LNPE-----GNVGDPMLDPALEAPADVPQVEKLEKFTASYGLCLD-----SVIFNMD 506
DB 4320 INKEIYVKNANQKNIDVDLGLQINPKIAHLNRTAKFKSTNNDIETNVINGSLVND 4379
QY 507 GTIELR 512
DB 4380 GKTSIR 4385
RESULT 7
IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
C:Species: Haemophilus influenzae
A:Variety: strain HK613
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C:Accession: C41859
R:Poulsen, K.; Reinholdt, J.P. Kilian, M.
J. Bacteriol. 174, 2913-2921, 1992
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
A:Reference number: A41859; MUID:92234949; PMID:1373717
A:Accession: C41859
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1849 <PO>
A:Experimental source: strain HK613
A:Note: sequence extracted from NCBI backbone (NCBIP:97285)
C:Superfamily: IGA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase
Query Match 5.4%; Score 149; DB 2; Length 1849;
Best Local Similarity 20.1%; Pred. No. 2.2;
Matches 117; Conservative 73; Mismatches 215; Indels 178; Gaps 26;
QY 28 VNKEKNAILYPHGDHHPIDHKPVGIG-----HSHSNVELFKPEGVAKK----- 75
DB 705 LVNEKTLFLSGRTPHARDI-----AGISSYKDPHTENNEVVVDDWVNNFKATIM 759
QY 76 --EGN-KVYTGELTNVLLNNTFNNOFTLANGQKRVSFSPPELEKLGIGIMLVK- 131
DB 760 NVTGNASLYSGR---NVANITSNITASN-----NAQVHIGY-----KTGDTVCVRS 802
QY 132 -----LITPDGKLVLEKSGKVGEGV-GNI-----ANFELDQYLPFGOTFKYTIASKDYP 180

803 DYTGYVTCHNSNLSEKALNSFNPTNLRGNVNLTENASFTLGRANLFG-----TIOSIGTS 857
QY 181 EVSYDCTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDLVRVDFSEFHGNA 240
DB 858 QVN-----LKENSHWHLTGSNNVNOLNLTNGHILHNAQNDANKVTTYNTLVNSLSNG 911
QY 241 -----YLENNYKVEIKLPI-----PKLNQGT-----TRTAGNKIPVTFM 275
DB 912 SFYYVDFTNKSNKVVVANKSATGNFTLQVADKTGEPNEHNETLFDASNATRNLEVT-L 970
QY 276 ANAYLDN-----QSTYIVVEPILKEKNOT-----DKPSILPOPKRN 311
DB 971 ANGSVORGAWKYKLRNVNGRYDLYNPEVKRNOTVDTTNTTTPNDIQADAPS----- 1022
QY 312 KAQENSKLDEKVEEP-----KTSKEVEKEKJSE 339
DB 1023 -AOSNNEEIAARVETPVPPAPAPATESAIASEQPETRPAETAQAPEMETNTANSTETAPKSD 1081
QY 340 TGNSTNSITLEEPTVDPVOEKVAKFAESYGMKLENVLENMDGTIELYLPSEGVKIKKNA 399
DB 1082 TATOTENPNSVESPS-----ETTEKVAENPPQENETVAKNEQATEPTFQNGEVAKEQDP 1136
QY 400 DF-----TGEAPOGNGENKPSNGKYSTGTVENQPT-----ENKPADSLPEAPNEKPY- 447
DB 1137 TVANTQTNEATOSEGKTETQ-----TAEKSEPTESVTSNQPEKTVSQSTEDKVVV 1191
QY 448 ----KPNSTONGMLNPEGVNGSDP-MLDPALEEAFAVDVQVE 485
DB 1192 EKBEKARVETEETQKAPQVTSKEPPKQAEPAPEVETDTNAEE 1234
RESULT 8
T18372
repeat organellar protein - Plasmodium chabaudi
C:Species: Plasmodium chabaudi
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted
A:Reference number: T18922; MUID:98418765; PMID:9747969
A:Accession: T18372
A:Molecule type: DNA
A:Status: preliminary; translated from GB/EMBL/DDRJ
A:Residues: 1-1939 <WER>
A:Cross-references: EMBL:043145; NID:g1151157; PID:g1151158; PIDN:AAC63403.1
Query Match 5.4%; Score 148; DB 2; Length 1939;
Best Local Similarity 19.3%; Pred. No. 2.7;
Matches 117; Conservative 101; Mismatches 206; Indels 182; Gaps 27;
QY 4 LDKKIEEKIAGIMKQYGVKRESIVWNKEKNAIYPHGDHHP-----ADPIDHKPV----- 54
DB 989 LGEHKVEWAGIEEKY--KVEATKLAERKDVYTKLGEQKKEIAKLEDGKHEVNVVEK 1046
QY 55 -----GIGHSHSN-----YELFKPEGVAKKGGKQKVTGTEELTNV 90
DB 1047 KNASLLNMLENHNKEMIKLKEEHKESASDLVEKLYQKDEEV-KNSNNKI---EELTNVI 1102
QY 91 NLLKNSFTFNQNFTRANQKRVSFSPPELEK-KLGINMLVKLITPDGKLVKESVG-KVF 148
DB 1103 KDLNDSIMCYKKQILEEVEKRNENY--FEINKLIVONNEKMDM--NDKKLEKENEIKKL 1158
QY 149 GEGVGNIANFELDQYLPFGOTFKYTIASKDYPEVSDGTFTVPTSLAYKMASOTIFYPFH 208
DB 1159 NKLSNKKVFPETK-----NTYK----- 1176
QY 209 AGDYLRVNPQFAVPKGTDALV--RVDEFHGNAYLENNYKVGEEK--LPIPKLNOQTTR 264
DB 1177 -----NSMWYNNENKRIYDVSCKENISESDVGG--KGGMLKMTLSKKERNIFS 1226
QY 265 TAGNKIPVTFMANAYLNOQSTYIVVEPILKEKNOTDKPSI-----LPQF 308

Db 1227 INDNRNSELVDTI---KSAVINKIEMKYKKEIDNGKNIEDLKNKILDLINLENM 1283
QY 309 KRKAQENSKL-----DEKVEEPTSEKVE-----KELSETGNSTNSYLE 350
Db 1284 KWLVDENNKLKETEIKDNKLNKEKNEITEILNLDIILKKEISEWKEDEEKLTK 1343
QY 351 EYPTVDVPQEVAKFAESYGMKLENVL--FN-----MDGTIELYLPSEGVKKNMAD 400
Db 1344 NI-----KLANDIEQINKVEKIEEENLMKFNENINEVSLKNQIEKMKLELNKVEL 1399
QY 401 FTGEAPOGENKPKSPENGKVSIGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNP 460
Db 1400 LLAERKRETNMSINDNKIVENILED--TDSK-----QNNLNK---NV 1438
QY 461 EGNVSDPMLDPALEAPAVDPQVKLEKFTASYGLGL-----DSVIFNMDG---TIEL 511
Db 1439 EDKTGDINCERNNOQAREISVLKDEIKISMLYGEELNKRKNSYDEKVKNTNELKEIKI 1498
QY 512 RLPSE 517
Db 1499 RNKKG 1504

RESULT 9
T39924
hypothetical protein SPAC21B10.03c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39924
R:McDougall, R.C.; Rajadream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, October 1997
A:Reference number: 221891
A:Accession: T39924
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-791 <MCD>
A:Cross-references: EMBL:AL121794; PIDN:CAB57927.1; GSPDB:GN00067; SPDB:SPAC21B10.03c
A:Experimental source: strain 972h; cosmid c21B10
C:Genetics:
A:Gene: SPAC21B10.03c; SPDB:SPAC21B10.03c
A:Map position: 2

Query Match 5.3%; Score 146.5; DB 2; Length 791;
Best Local Similarity 22.1%; Pred. No. 0.86;
Matches 115; Conservative 67; Mismatches 193; Indels 155; Gaps 26;

QY 46 DPIDHKPV-GIGHSHSNVELFKPEGVAKKGNKVYTGEEELTNVNLKLNSTNNONFT 104
Db 126 DPEDAGVPLAGL-----EESTDNVEWDQFATNEKLFV-----KSHFEDELYT 168
QY 105 LANGOKRVSFSPPELEKKLGINMLVK---LITPDGKLVKSVKSGVGEV-----G 153
Db 169 -----SRIDRSHPKYKEQEADRIAKIEGTVNNIHAERGLKVDSDGLDEEDLYSG 223
QY 154 NTANFELDQPLPGQTFKYITASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTY 213
Db 224 VERSIDVVRNYT--RSNAYNNKNDQKPKNEAPHQHP-----QKVVYPP---DDPA 270
QY 214 LRVNQFAVPK--GTD--ALVRVDFPHGNAYLENNYKVGSIK-----LPTKLNQ 261
Db 271 IVSHRLALPRAGDPSRAAERFFNARKAGPLSRREKEGQIKFQMPSQSLKIGSLD-- 328
QY 262 TTRTAGNIPVTFMANAYLDNQSTIVIVETILEKENQNDKPSILPQFKRKAQENSKLDE 321
Db 329 -----SKQP-----SSTKSAEVKVADEKQLPDASS-----QATADSKPEKPE 367
QY 322 KYEPEKT-----SKVKEKLESETGNSTNSLTLEEVTVDFVQEK-VAKFAESYGMKL- 373
Db 368 EAEKPVTSATEVSSEKVEK---VDGNTSSPSKEEKEKPEKPSVVTQRKETTGTGKLG 423
QY 374 -----ENVLFNMDGTIELYLPSEGVKKNMADFTGEAPQ----- 407
Db 424 TKLNKAKISFRPNVAAPVETGKFTIPSKPAPVNASRPMQPSNNEASIPSTTPQSPS 483

QY 408 --GNGENKPS-----NGKVST-----GTVENOPTENKPAD-----SLPEAPNEK 445
Db 484 VVNGENKPSPPYFFNGPVSSSEKEPILDNFNVFNKNGEEHQAEQIDKPFSCPTWNTIG 543
QY 446 PVKPNSTNGMLN--PEGNVSDPMLDPALEAPAVDPV 483
Db 544 P-----NSLQQTIANRPEGNGSGS-----AKKAAANPM 572

RESULT 10
C90593
hypothetical protein MYPV_6510 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: C90593
Nucleic Acids Res. 29, 2145-2153, 2001
A:Chamcaud, I.; Heiliger, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: C90593
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1272 <KUR>
A:Cross-references: GB:AL445566; PID:g14090066; PIDN:CAC13824.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPV_6510
A:Genetic code: SGC3

Query Match 5.3%; Score 146; DB 2; Length 1272;
Best Local Similarity 20.3%; Pred. No. 1.9;
Matches 125; Conservative 94; Mismatches 238; Indels 158; Gaps 28;

QY 1 MKDLDDKIEEKIAGIMKOYGVKRESIVVNEKNAILIYPHGDHHDHAFIDEHKVPVGTGSH 60
Db 86 LTDTDTQDEKINVSX-----LISNEQK-----QEENSQRNSQN 121
QY 61 SNYELFKPEGVAKKGNKVYTGEEELTNV-----NLLKNSTFNQNFNFTLANGQRVSV 114
Db 122 QNEKFANDDEGTLKVLVEFSYNGVPTREVKLHGFKNTIQKTSNNLNLNLSNKEQ--- 177
QY 115 SFPPEL---EKKLGINMLVK--LITPDGKLVKSVKSGVGEVGNIANFELQPYLPQG 168
Db 178 IHPSELDSNQQLVTFNFKLSLETLDIEKYVLYEVYVNDIDAEAKIRLTLQKLTNE 237
QY 169 TFK--YTIASKDYPEVSYDGT-----TVPTSLAYKMASQTIFFPHAGDTYLRVNP-- 218
Db 238 NIKNSFEVLKDFKYNLDSDSLKAKINLVKNDLSIEELLNITSKGFANKTDDOYNKLG 297
QY 219 --QFVAPKGTDALVRVDFDEFGHAYLENNYKVGEL--KLPIKLNQGTTRTAGNIPVT 273
Db 298 ETLNIDLPNGYEF-----EFVSLAPKANDASVGLTYKLVKNLNDNGTSENFSK---- 347
QY 274 FMANAYLDNQSTIVIVEP--ILEKENQTDKPSILPQFKRKAQENSKLDEKVEPEKTSK 331
Db 348 --NGRIET--NTIVEETNLSLKNDSKDEP-----KKDNSNTNSKDEPKTEDEPKVEEP 397
QY 332 VEKE--KLSETGNSTNSLTLE-----VPTVDVQEKVA----- 363
Db 398 REDEPKTNPSDSKDKPKIDPENEKPKDQPKTEPKNEPKDKPKVEPKDETLAIFDKISKI 457
QY 364 -----KFAESYGMKLEN--VLFNMDGT--IELYLPSEGVKKNMADTGEA 405
Db 458 ELKENSOLKQKLSQFKRES--DLNLSNLKVLVSODKNNKFSLSLPOGYSTISFKLA----- 510
QY 406 PQNGENKPSNGKVSSTGV-----ENQPTENKPADSLPEAPNEKPVKPNSTNGMLN 459
Db 511 -----SNSNDEGLDVKVIVQKGEVKKELKLTLLTFEFSLK-----ESDQDLDFS 561
QY 460 PEGNVSDPMLDPALEAPAVDPVQEK-----LEKFTASYGL--GLDSVIFNMDGTIELR 512
Db 562 KKDPLASSVWNDDKIITRESLV--VKNKTIENFDNFNVDISYSVSSSDEV----NGLKIK 615

B86807
hypothetical protein yoiC [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: B86807
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: B86807
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1441 <STO>
A:Cross-references: GB:AE005176; PID:g12724450; PIDN:AAK05556.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yoiC

Query Match 5.2%; Score 141.5; DB 2; Length 1441;
Best Local Similarity 22.9%; Pred. No. 4.1;
Matches 113; Conservative 52; Mismatches 185; Indels 143; Gaps 29;

Qy 82 TGEELTNVYLLKSTFNQNTFLANGQKRVSFSPPELEKKGILNMLVLIPT-DGK-- 138
Db 955 TGASISSVANITVNDSSNTINPTDGTGGNTT-----NPTDGNIT 994

Qy 139 VLEKSGKVFEGVGNITANFELDPYLPQTFKTYIASK-----DYPEVSYD---GTFETVP 191
Db 995 VLPQDQGLK---GTADDVTVTDQPLSPGSDGVTLPDQKVRDPGSGVNPVGTVDP 1051

Qy 192 TSLAYKMASQIFYPFHAGDTYLRVNPQFVAPKGTDALVRVEDEFGHNAVLENNYKV--- 248
Db 1052 DG-----TIHLP---GGVINGSGSVTP-GPDGKTGTDGDD---TTLNPNSPVVG 1095

Qy 249 --GEIKLPIPKLNOGTRTA-GNKIPVTVMANAYLDNOSTYIVVEVILEKENQTDKPSIL 305
Db 1096 DNGSVTLT---GGGTASTPNGN---ITLPGTGVDPDGT---IHLPGGDIVNPDGTTILP 1146

Qy 306 PQKRNKAQENSKLDEKVEEPEKTSKVEKEKLSSTNSTLEEVP---TVDPPQCEK 362
Db 1147 QDQKGTGTDGDKI--RPNGFIIDGNSVTLPGGGVITPVGGINVPGSGVDP----- 1199

Qy 363 AKFAESYGMKLENVFNMDGTIELYLSGSEVKNAMADFTGEAPQGNKPSNGKSVST 422
Db 1200 -----DGTV--HLPGGDIVN---PDGTTILPGQDGTGTDGDKVK- 1235

Qy 423 GTVENOPTENKPADS--LP-----EAPNEKPVKPNST---DNGMLNPEGNV- 464
Db 1236 ---PNGFSINPDGSIITLPGGGVITPVGGINVPGSGVDPDGTVHLPGGGVWNPDGTTIT 1292

Qy 465 -----GSDPMLDPALPEEAPVD-----PVQEKLEKFTASYGLDLSVFNMD 506
Db 1293 LPQDQKGTGTDGDDVNIKP---NGFSVSNPDGSIITLPGGGVITNPGGTYNVPGGTVV-DPD 1348

Qy 507 GTIELPLPSGSEVI- 519
Db 1349 GTI--HLPNGSEVI 1359

RESULT 14
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
e.
A:Reference number: Z18847; MUID:93209931; PMID:8458832
A:Accession: T18262
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-1664 <FUJ>
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match 5.2%; Score 141.5; DB 2; Length 1664;
Best Local Similarity 21.7%; Pred. No. 5;
Matches 86; Conservative 54; Mismatches 156; Indels 101; Gaps 17;

Qy 117 PPELEKKGILNMLVLIPTDQKLVKSVKVFEGEGVGNIANFELDPYLPQGTFFKTYIAS 176
Db 597 PSELPSYVIMELDKTKVREGDV---IATIRVNNIKNLGAGIQIGIKYDPKYLEANIET 653

Qy 177 KD-----YPEVSYDGTFT-----VETSLAYKMASTIETIYPPHAGDTYLRVNPQFAPVKG 225
Db 654 GDPIDEGTTPAVG--GTILKNRDYLTGTVAINNVSXGIL-----NFA----- 693

Qy 226 TDALVRVDEFGHNAVLENNYKVGEEKLPIPKLNOGTTR-----TAGNKIPVTVMANAYL 280
Db 694 --AYYVFDYREEGKSEDTGIIGNTGFRVLKAEITIRFEELSPGSDGTIYMLDWYL 751

Qy 281 DNSTIV--EVPILKQENQ---TDKPS--ILPQKRNKAQENSKLDEKV--EETKISEK 331
Db 752 NRISGVYVIQAPAIKAASDEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 811

Qy 332 VEKEKLSSTNSTNSTLEEVTVDQKVAFAESYGMKLENVFNMDGTIELYLP 391
Db 812 PEPITDTPSDPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 841

Qy 392 EVIKKNMADFTGEAPQGNKPSNGK--VSTGTVENOPT-ENKPADSLPEAPNEKPYK 448
Db 842 E-----TPSDEPTSETPEEPIPTDTSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 888

Qy 449 PENSTNGMLNPGNSGSDPMLD---PALPEAPVDP 482
Db 889 SDEPTPSE--TPEEPIPTDTSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 923

RESULT 15
A60234
Iga Fc receptor precursor - Streptococcus agalactiae (strain SB35)
N:Alternate names: Iga-binding protein; protein Bac
N:Contains: beta antigen
C:Species: Streptococcus agalactiae
C:Date: 08-Dec-1992 #sequence_revision 08-Dec-1992 #text_change 26-Aug-1999
C:Accession: A60234; S14595; A60230
R:Heden, L.O.; Frithz, E.; Lindahl, G.
Eur. J. Immunol. 21, 1481-1490, 1991
A:Title: Molecular characterization of an Iga receptor from group B streptococci: seq
ents with Iga-binding capacity.
A:Reference number: A60234; MUID:91257158; PMID:2044657
A:Accession: A60234
A:Molecule type: DNA
A:Residues: 1-1134 <HED>
A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CAA41384.1; PID:g46521
A:Note: the source is designated as group B streptococcus strain SB35
R:Heden, L.; Frithz, E.; Lindahl, G.
Submitted to the EMBL Data Library, March 1991
A:Description: Molecular characterization of an Iga receptor from group B streptococ
fragments.
A:Reference number: S14595
A:Accession: S14595
A:Molecule type: DNA
A:Residues: 1-1134 <HE2>
A:Cross-references: EMBL:X58470; NID:g46520; PIDN:CAA41384.1; PID:g46521
A:Note: the source is designated as Streptococcus agalactiae
R:Lindahl, G.; Akerstroem, B.; Vaerman, J.P.; Stenberg, L.
Eur. J. Immunol. 20, 2241-2247, 1990
A:Title: Characterization of an Iga receptor from group B streptococci: specificity f
A:Reference number: A60230; MUID:91055397; PMID:2242758
A:Accession: A60230
A:Molecule type: protein
A:Residues: 'X', 39-48, 'X', 50-52, 'X', 54-56 <LIN>
C:Superfamily: Iga Fc receptor
C:Keywords: Cell wall; immunoglobulin receptor; tandem repeat; transmembrane protein

F:1-37/Domain: signal sequence #status predicted <SIG>	
F:138-1134/Product: IGA fc receptor #status experimental <MAT>	
F:199-438/Domain: IGA binding #status predicted <IGAI>	
F:439-826/Domain: IGA binding #status predicted <IGAI>	
F:827-915/Region: proline-rich repeats	
F:916-1101/Domain: cell wall-spanning #status predicted <CWS>	
F:1102-1129/Domain: transmembrane #status predicted <TM>	
Query Match 5.1%; Score 141; DB 2; Length 1134;	
Best Local Similarity 20.0%; Pred. No. 3.1;	
Matches 126; Conservative 95; Mismatches 207; Indels 202; Gaps 32;	
Qy 2	KDLKKIEKI-----AGLMK-QYGVKRESIVNKKENAIYPHGDHHDPI--- 48
Db 240	EDAEVKVREELGKLFSSTKAGLDQEIQHVAKET---SSEENT---QKVDEHYANSLQNL 293
Qy 49	-----DEHKVGVGHSHSNVELFKPEGVAKKEGNKY--TGEELTNVNVLL 93
Db 294	AOKSLEELDQATTNEQATQVKNQNFLENAQKLKEIQPLIKETNVKLYRAMSESLQVEKEL 353
Qy 94	KNSTENNQNFILANGOKRVSTSFPELEKKGILNMLVKLIITPDGKVLKFSKVGVGEGVG 153
Db 354	KHNSANLEDLVAKSEIVR-----EYEGKL--NOSKNL--PELKQLEEEAHSKLQWVE 404
Qy 154	NT-ANFELDQPYLPQOTFKYTIASKD-----YPE--VSYDG---TFTVPTSLEYK 197
Db 405	DFRKKFKTSEQVTPKKRYKRDLAANNQKQIETLTVSPENITVYEGEDVKFTVTA---K 460
Qy 198	MASQIFIPFFHAGDYLRVNPQFVAPKGTDALRVDFDEFGHGNAYLENNYKYGEIKLPIPK 257
Db 461	SDSKT---TLDFSDLLATKYNPSV-----DRISTNYKNT---DNHKIAETITKLNK 506
Qy 258	LNQGTTRT-----AGNKIPVTFMANAYLDNQSTYIVVEPILEKENOTDKPSILPQPKRN 311
Db 507	LNESQTVTLKAKDDSGNVVETK-----TITVQKREEK-----QVEKT 544
Qy 312	KAOENSKLDEKV-EEPKTSKVEKEKLESETGN-----STNSLTLEEVPTV----- 355
Db 545	PEQKDSKTEEKYPQEPKSNKNDKNQLOELIKSAQOELEKLEKAIKELMEQPEIPSNPEYGIQ 604
Qy 356	-----DPVQEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSGEVIK 395
Db 605	KSINWSQKEPIQEAITSFKKIIGDSSSKYYTEHYFNKYKSDFMNVLHAQM-----EMLT 659
Qy 396	KNNADFTGEAPQNGENKPSNGKVST-----GTVENOPT-----NK----- 433
Db 660	RKVQVMNKPDPNAETIKKIFESDMKRTKEDNYGSLNDALKGYFEKYEFTLPFNKIKQIVD 719
Qy 434	PADSLPEAPNEKPKVPKPNSTONGMLNPEGNVGSDFMLDPALEEA-----PAVDPVQ 484
Db 720	DLDKVQEQDQAPI-PENSE-----MQAKEKAKIAVSKYKSVLDGVH 762
Qy 485	EKLEKFTASYGLGL-----DSVIFNMD 506
Db 763	QHLQKKNHSKIIVDLFKLELAIKQQTIFDID 792

Search completed: May 13, 2003, 13:57:16
Job time : 25.9447 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:50:17 ; Search time 6.19355 seconds
(without alignments)
3535.857 Million cell updates/sec

Title: US-09-471-255-10

Perfect score: 2746

Sequence: 1 MKDLKKIEIKIMQYQ.....IELRLPSGEVKKNLSDFIA 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149	5.4	1849	IGA4_HAEIN	P45386 haemophilus
2	143.5	5.2	1139	HMW1_MYCGE	Q09413 mycoplasma
3	141.5	5.2	1664	SLP1_CLOTPM	Q06852 clostridium
4	141	5.1	1164	BAG_STRAG	P27951 streptococc
5	140	5.1	1013	SCA4_RICRH	Q9aj81 rickettsia
6	137.5	5.0	1526	TP2A_CRIGR	P41515 cricetus
7	134	4.9	357	TP2B_RAT	P19814 rattus norv
8	133	4.8	710	LP78_LARATH	Q06738 arabidopsis
9	132.5	4.8	857	NFM_CHICK	P16053 gallus gall
10	131.5	4.8	1701	MSPI_PLAFM	P08569 plasmodium
11	131	4.8	774	GLYB_PLAFG	P02895 plasmodium
12	130.5	4.8	1183	CNA_STAAU	Q33654 staphylococ
13	130.5	4.8	1750	Y832_METJA	Q58242 methanococc
14	130	4.7	1702	IGA2_HAEIN	P45384 haemophilus
15	128.5	4.7	366	YMR3_YEAST	P33885 saccharomyc
16	128	4.7	1275	YAU9_SCHPO	Q10164 schizosacch
17	127.5	4.6	1612	TP2B_MOUSE	Q64511 mus musculu
18	127	4.6	1478	BK11_YEAST	Q01389 saccharomyc
19	127	4.6	1612	TP2B_CRILLO	Q64399 cricetus
20	126.5	4.6	1167	SCA1_STRPY	P15926 streptococc
21	126.5	4.6	1701	MSPI_PLAFW	P13819 plasmodium
22	126	4.6	515	YMR1_YEAST	P35741 saccharomyc
23	126	4.6	537	TEEG_STRPY	P18481 streptococc
24	126	4.6	682	GR78_YEAST	P16474 saccharomyc
25	125	4.6	1694	IGA0_HAEIN	P44969 haemophilus
26	125	4.6	818	PK2_YEAST	P47116 saccharomyc
27	125	4.6	1012	SCA4_RICSL	Q9aj80 rickettsia
28	123	4.5	725	ADCB_MOUSE	Q9qy88 mus musculu
29	123	4.5	1076	NUPI_YEAST	P20676 saccharomyc
30	123	4.5	1630	MSPI_PLAFK	P04932 plasmodium
31	123	4.5	1637	MRSP_STAAU	P80544 staphylococ
32	123	4.5	1639	MSPI_PLAFW	P04933 plasmodium
33	122.5	4.5	547	CATX_BACSU	P94377 bacillus su

ALIGNMENTS

RESULT 1

ID	IGA4_HAEIN	STANDARD	PRT	1849 AA
AC	P45386;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).			
GN	IGA.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
OC	Haemophilus.			
OX	NCBI_TaxID=727;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NHIT HK61;			
RX	MEDLINE=92234949; PubMed=1373717;			
RA	Poulsen K., Reinholdt J., Kilian M.;			
RT	"A comparative genetic study of serologically distinct Haemophilus influenzae type 1 immunoglobulin A1 proteases.";			
RL	J. Bacteriol. 174:2913-2921(1992).			
CC	-!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A			
CC	PRODUCING INTACT PC AND FAB FRAGMENTS.			
CC	-!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at			
CC	certain Pro- -Xaa bonds in the hinge region. No small molecule			
CC	substrates are known.			
CC	-!- SUBCELLULAR LOCATION: Secreted.			
CC	-!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC			
CC	SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE			
CC	OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE			
CC	DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY			
CC	SIMILARITY).			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	M87491; AAA24968.1; ..			
DR	MEROPS; S06.001; ..			
DR	InterPro; IPR000710; Iga_S6.			
DR	InterPro; IPR004899; Pertactin_sup.			
DR	Pfam; PF023395; IGA1; 1.			
DR	Pfam; PF03212; Pertactin; 2.			
DR	PRINTS; PR00921; IGASERPTASE.			
DR	Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.			
FT	SIGNAL	1	25	POTENTIAL.
FT	CHAIN	26	1021	IMMUNOGLOBULIN A1 PROTEASE.
FT	PROPEP	1022	1849	HELPER PEPTIDE (POTENTIAL).
FT	ACT_SITE	299	299	PROBABLE.
SQ	SEQUENCE	1849 AA;	202957 MW;	79A7D018C7150AEA CRC64;

Query Match 5.4%; Score 149; DB 1; Length 1849;
 Best Local Similarity 20.1%; Pred. No. 1;
 Matches 117; Conservative 73; Mismatches 215; Indels 178; Gaps 26;

QY 28 VNEKNAIIPVHGDDHAPDIHDKPVGIG-----HSHSNVLEKPEGVAKK----- 75
 DB 705 LNVKGTGLSGRTPHARDI-----AGISSPKDPHTFENNEVVVDDWIHNRKATIM 759
 QY 76 --EGN-KVYTGEELTNVALLKNTNNFTLANGQKRVSFSPPELEKKGINMLVK- 131
 DB 760 NVITGNASLYSGR---NVANITSNITASN---NAQVHIGY-----KTGDTVCVRS 802
 QY 132 ----LITPGKVLKRVSGVFGEV-GNI-----ANFELDQYLPQGTFKYTIASKDVP 180
 DB 803 DYGYVYCHNSNLSKALNSFNPTNLRGVNLTENASFTLGRANLFG-----TISIGTS 857
 QY 181 EVDYDGTFTVPTSLAYKMASQTIYFFHAGDTYLRVNPQFAPVPGTDALRVDFDEHNA 240
 DB 858 QVN-----LKENSHWLTGNSNVNGLNTNGHILHNAQNDANKVTTYNTLVNSLSGNG 911
 QY 241 -----VLENNYKVGEEKLPI-----PKLNOGT---TRTAGNKIPVTFM 275
 DB 912 SFYVWDFTNKSNKVVVKNKATCNFTLOVADKTCGPNHNLTPDASNATRNLEVT-L 970
 QY 276 ANAYLN-----QSTYIVVEVPILEKENOT-----DKPSILPOFKRN 311
 DB 971 ANGSVDGANKYKLRNVNGRYDLYNPEVEKRNQTVDTNITTPNDIQADAPS----- 1022
 QY 312 KAQENSKLDSEKVEP-----KTSSEKVEKELSE 339
 DB 1023 -AOSNNEEIARVETVPVPPAPAPATESIASEQPTRPAETAQAQMBETNTANETAPKSD 1081
 QY 340 TGASTNSTLEEVTDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKKMA 399
 DB 1082 TAOQENPSESVPs-----ETTEKVAENPPQENETVAKNEQEATEPTQNCSEVAKEDQP 1136
 QY 400 DF-----TGAPAGNENKPSNGKSTGTVENQPT-----ENKPADSLPAPNEKPV- 447
 DB 1137 TVEANTQNEATQSEKTEETO-----TAEIKSEPTSEVTSENQPEKTVSQSTEDKVVV 1191
 QY 448 ----KPNSTDNGLNPNEGNVGSDP-MLDPALDEEAFVDPVOE 485
 DB 1192 EKEKAKVETEETQKAPQVTSKEPPKQAPAEVETDINAEE 1234

RESULT 2
 HMW1_MYCGE STANDARD; PRT; 1139 AA.
 AC 049413; 049365;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytadherence high molecular weight protein 1 (Cytadherence accessory
 protein 1).
 DE HMW1 OR MG312.
 GN Mycoplasma genitalium
 OS Mycoplasma genitalium
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 CX NCBI_TaxID=2097;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 Fleischmann R.D., Bult C.J., Kierlavage A.R., Sutton G., Kelley J.M.,
 Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
 Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
 Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RA "The minimal gene complement of Mycoplasma genitalium".
 RT Science 270:397-403(1995).
 RL [2]
 RN SEQUENCE OF 721-847 FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;

RX MEDLINE=94075230; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
 RT "A survey of the Mycoplasma genitalium genome by using random
 RT sequencing.";
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -!- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
 CC STABILIZES THE SHAPE OF THE WALL-LESS MYOPLASMA. THIS
 CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
 CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERIN PROTEINS
 CC IN THE MYOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
 CC MEMBRANE (BY SIMILARITY).
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U39712; AAC71534.1; -;
 CC EMBL: U02261; RAD12527.1; -;
 CC TIGR: MG312; -;
 KW Cytadherence; Structural protein; Complete proteome.
 SQ SEQUENCE 1139 AA; 130531 MW; 0011D3288C3DD856 CRC64;

Query Match 5.2%; Score 143.5; DB 1; Length 1139;
 Best Local Similarity 20.8%; Pred. No. 1.1;
 Matches 128; Conservative 93; Mismatches 236; Indels 157; Gaps 33;

QY 26 IVNKEKNAIIPVHGDDHHH--ADPIDEH-KPVGICHSHSNVLEK----PEEGVAKKEGN 78
 DB 52 IAFNKETGVVYDPYDGTEDYDISQLFDENGPNFVDEKQENDYLYKRYVGNPDYGSYDNGE 111
 QY 79 KVTYGEELTNVALLKNTFNQNTLANGQKRVSFSPPE-----LEKGLGINMLVKL 132
 DB 112 WWSGYFENDQWISTKESQPTDENYGFDS-----DLPEVKQPEVEDNYGFD--ND 161
 QY 133 ITPDGKVLKRVSGKVGEGVGNIANFELDQYLPQGTFKYTIASKDY----- 179
 DB 162 LPPEVKQPEVEDNY--GENDLPPPEVKQP-----ESVVDQPSDDYFAKQPTDENYGF 214
 QY 180 ----PEVSY-DGTFVPTS---LAYKMASQTIYFFHAGDTYLRVNPQFAPVPGTDALVR 231
 DB 215 NDLPEVKQPEVVDQPSDDHFAKQPESTTDSYSF---DSDL-PQPTLDQPSLDDHVVQY 270
 QY 232 VFD---EFHGNAYLENNYKVG--EIKLPKLNQOQTTTAGNKIPVTFMANAYLDNGSY 286
 DB 271 NFDHHEELKPVAEHONNYOVGFQVQANLDN-NEEIQPTAEKKVTTDF-----ESKQAO 323
 QY 287 IV---EVPV-LEKENQT-----DKPSILPOFKRNKAQENSKLDEKV-EEPKTSKVEK 334
 DB 324 VVDSYQLPIDTDQDQDTTFSSSFETQPTV-EQFDQVNSVNDQFKPEITKEPVLESFNK 382
 QY 335 EKLSTGTGSTNSTL-----EEVPTVDPVQEKVAKFAESY- 369
 DB 383 QDVVETSLNSESNNLYSENKNDATNNDSENSEFQLNLSNSETASDDVHYESKSEPIHYK 442
 QY 370 -----GMKLENVLFNMDGTIELYLPs-GEVTKNNMADFTGEAPGNENKPS 416
 DB 443 FGSDLQSQSNNSNLSSEPVKFNSETAPDAHESQSEPVQVQYDI-----YQNEELKPTL 497
 QY 417 NGKVTGTGTVENQPTENK-----PADSLPEAPNEK-----VKPENSTDMGLNPE 461
 DB 498 DQPSDDYFAKQPTDENYGFNDLPPEVKQPEVVDQPSDDHFAKQPESTTDSISDSD 557
 QY 462 GNVGSDPMLDPALEAPAVDPVQ-----EKLEKFTA---SYGLGLSVIENMDGTIELR 512
 DB 558 -----LPQPTLDQPSLDDHVVQYNFHDHHEELKPVAEHONNYQVGFQVQANLDNNEIQ 610
 QY 513 LPSGEVTKNLSDF 526

Db 611 -PIAE--KEVTFD 621

RESULT 3

SLP1_CLOTH STANDARD; PRT; 1664 AA.

AC Q06852; 1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).

DE OLPB.

GN OLPB.

OS Clostridium thermocellum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1515;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCIB 10682;

RX MEDLINE=93209931; PubMed=8458832;

RA Fujino T., Beguin P., Aubert J.-P.;

RT "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CtpA and a protein possibly involved in attachment of the cellulosome to the cell surface.";

RL J. Bacteriol. 175:1891-1899(1993).

CC -! SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.

CC -! SUBCELLULAR LOCATION: Cell wall.

CC -! SIMILARITY: CONTAINS 4 S-LAYER HOMOMOLOGY (SLH) DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL: X67506; CAA47841.1;

CC InterPro: IPR001119; SLH.

CC Pfam: PF00395; SLH: 3.

CC PROSITE: PS01072; SLH_DOMAIN: 2.

CC Cell wall; S-layer; Signal; Repeat.

CC SIGNAL 1 28 POTENTIAL.

CC CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.

CC DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.

CC REPEAT 36 191 1.

CC REPEAT 207 363 2.

CC REPEAT 409 565 3.

CC REPEAT 607 763 4.

CC DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF T-P-S-D-E-P.

CC DOMAIN 1378 1449 GLY/PRO/SER/THR-RICH.

CC DOMAIN 1453 1494 SLH 1 (INCOMPLETE).

CC DOMAIN 1495 1565 SLH 2.

CC DOMAIN 1566 1825 SLH 3.

CC DOMAIN 1626 1646 SLH 4 (INCOMPLETE).

CC SEQUENCE 1664 AA; 171819 MW; 5F396695BA9FE74B CRC64;

Query Match

Best Local Similarity 5.2%; Score 141.5; DB 1; Length 1664;

Matches 86; Conservative 21.7%; Pred. No. 2.3;

17: Indels 101; Gaps 17;

QY 117 PBELEKLGINMLVKLITPDGKVKLEKVGKVGEGVGNIANFELDQYLPQOTFKYTIAS 176

Db 597 PSELPSDYVIMELDKTKVKEGDV---IIATIRVNKNKLAGYQIGIKYDPKVLFAFNIE 653

QY 177 KD-----VPEVSYDGTFT-----VPTSLAYKMASOTTFEYFHAGDYLRVNPQFAVPKG 225

Db 654 GDPIDEGTWPAG--GTLIKNDYLIPTGVAINNVSGLI-----NFA----- 693

Y 226 TDALRVFDFEFGNAYLENNYKVGKIKLPIKLNQGTTR-----TAGNKKIPVTMANAYL 280

Db 694 --AYVYFDYREBKGSEDGTGIIGNIGFRVLKAEATTIRFELESMPSGIDGTYMLDWYL 751

QY 281 DNQSTYIV--EVPILEKENQ---TDKPS--ILPQFKRNKAQENSKLDEKV--EKPKTSEK 331

Db 752 NRISGVVVIQAPAKAASDEPIPTDPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSET 811

QY 332 VEKEKLSETGNSNSTLEEVPTVDFOEKVAKFAESYGMKLENVLFNMDGTIELYLP 391

Db 812 PEEPIPTDPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPD 841

QY 392 EVIKNMADETCGAQOGNGENKPSNGK--VSTGTVENQPT-ENKPADSLPEAPNEKPVK 448

Db 842 E-----PTPSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 888

QY 449 PENSTNDGMLNEGVGSDPMLD---PALEAPAVDP 482

Db 889 SDEPTPSE--TPEEPIPTDPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 923

RESULT 4

BAG_STRAG STANDARD; PRT; 1164 AA.

AC P27951;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE IGA FC receptor precursor (Beta antigen) (B antigen).

GN BAG.

OS Streptococcus agalactiae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1311;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.

RC STRAIN=LA339;

RX MEDLINE=91312121; PubMed=1857207;

RA Jernstroom P.G., Chhatwal G.S., Timmis K.N.;

RT "The IGA-binding beta antigen of the c protein complex of Group B streptococci: sequence determination of its gene and detection of two binding regions";

RT Mol. Microbiol. 5:843-849(1991).

RN [2]

RP IDENTIFICATION OF IG-LIKE DOMAIN.

RX MEDLINE=97035265; PubMed=8880921;

RA Bateman A., Eddy S.R., Chothia C.;

RT "Members of the immunoglobulin superfamily in bacteria.";

RL Protein Sci. 5:1939-1942(1996).

CC -! SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (Potential).

CC -! SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL: X59771; CAA42442.1;

CC PIR: S15330; FCSOAG.

CC InterPro: IPR004829; Csurface_antigen.

CC InterPro: IPR001899; Gram_pos_anchor.

CC InterPro: IPR003599; Ig.

CC Pfam: PF00746; Gram_pos_anchor; 1.

CC ProDom: PD153432; Csurface_antigen; 2.

CC SMART: SM00409; Ig; 1.

CC TIGRFAMs: TIGR01167; LPXTG_anchor; 1.

CC TIGRFAMs: TIGR01168; YSTRK_signal; 1.

CC PROSITE: PS50847; GRAM_POS_ANCHORING; 1.

KW Cell wall; Peptidoglycan-anchor; Receptor; Signal;

KW Immunoglobulin domain.

Db 611 -PIAE--KEVTFD 621

RESULT 3

SLP1_CLOTH STANDARD; PRT; 1664 AA.

AC Q06852; 1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).

DE OLPB.

GN OLPB.

OS Clostridium thermocellum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1515;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCIB 10682;

RX MEDLINE=93209931; PubMed=8458832;

RA Fujino T., Beguin P., Aubert J.-P.;

RT "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein CtpA and a protein possibly involved in attachment of the cellulosome to the cell surface.";

RL J. Bacteriol. 175:1891-1899(1993).

CC -!- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.

CC -!- SUBCELLULAR LOCATION: Cell wall.

CC -!- SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL: X67506; CAA47841.1;

CC InterPro: IPR001119; SLH.

CC Pfam: PF00395; SLH: 3.

CC PROSITE: PS01072; SLH_DOMAIN: 2.

CC Cell wall; S-layer; Signal; Repeat.

CC SIGNAL: 1 28 POTENTIAL.

CC CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.

CC DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.

CC REPEAT 36 191 1.

CC REPEAT 207 363 2.

CC REPEAT 409 565 3.

CC REPEAT 607 763 4.

CC DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF T-P-S-D-E-P.

CC DOMAIN 1378 1449 GLY/PRO/SER/THR-RICH.

CC DOMAIN 1453 1494 SLH 1 (INCOMPLETE).

CC DOMAIN 1495 1565 SLH 2.

CC DOMAIN 1566 1825 SLH 3.

CC DOMAIN 1626 1646 SLH 4 (INCOMPLETE).

CC SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;

Query Match

Best Local Similarity 5.2%; Score 141.5; DB 1; Length 1664;

Matches 86; Conservative 21.7%; Pred. No. 2.3;

17: Indels 101; Gaps 17;

QY 117 PBELEKLGINMLVLIITPDGKVKLEKVGKGVGEGVGNIANFELDQYLPQOTFKYTIAS 176

Db 597 PSELPSYVIMELDKTKVKEGDV---IIATIRVNKNKLAGYQIGIKYDPKVLFAFNIE 653

QY 177 KD-----VPEVSYDGTFT-----VPTSLAYKMASQTIFFPHAGDYLRVNPQFAVPKG 225

Db 654 GDPIDEGTWPAG--GTLIKNDYLIPTGVAINNVSKGIL-----NFA----- 693

Y 226 TDALRVFDFEFGNAYLENNYKVGKIKLPIPKLNGGTTR-----TAGNKKIPVTMANAYL 280

Db 694 --AYVYFDYREBKGSEDGTGIIGNIGFRVLKAEATTIRFELESMPSGIDGTYMLDWYL 751

QY 281 DNQSTYIV--EVPILEKENQ---TDKPS--ILPQFKRNKAQENSKLDEKV--EKPKTSEK 331

Db 752 NRISGVVVIQPAPIKAASDEPIPIOTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSET 811

QY 332 VEKEKLSETGNSNSTLEEVPTVDFVOEKVAKFAESYGMKLENVLFNMDGTIELYLP 391

Db 812 PEEPIPTDTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPD 841

QY 392 EVIKNMADETCGAQOGNGENKPSNGK--VSTGTVENQPT-ENKPADSLPEAPNEKPVK 448

Db 842 E-----PTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPTSP 888

QY 449 PENSTNDGMLNEPVGSDPMLD---PALEAPAVDP 482

Db 889 SDEPTPSE--TPEEPIPTDTPSDEPTSPDEPTSPDEPTSPDEPTSPDEPTSPDEPT 923

RESULT 4

BAG_STRAG STANDARD; PRT; 1164 AA.

AC P27951;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE IGA FC receptor precursor (Beta antigen) (B antigen).

GN BAG.

OS Streptococcus agalactiae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1311;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.

RC STRAIN=LA339;

RX MEDLINE=91312121; PubMed=1857207;

RA Jerlstrom P.G., Chatwal G.S., Timmis K.N.;

RT "The IGA-binding beta antigen of the c protein complex of Group B streptococci: sequence determination of its gene and detection of two binding regions";

RT Mol. Microbiol. 5:843-849(1991).

RN [2]

RP IDENTIFICATION OF IG-LIKE DOMAIN.

RX MEDLINE=97035265; PubMed=8880921;

RA Bateman A., Eddy S.R., Chothia C.;

RT "Members of the immunoglobulin superfamily in bacteria.";

RL Protein Sci. 5:1939-1942(1996).

CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (Potential).

CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL: X59771; CAA42442.1;

CC PIR: S15330; FCSOAG.

CC InterPro: IPR004829; Csurface_antigen.

CC InterPro: IPR001899; Gram_pos_anchor.

CC InterPro: IPR003599; Ig.

CC Pfam: PF00746; Gram_pos_anchor; 1.

CC ProDom: PD153432; Csurface_antigen; 2.

CC SMART: SM00409; Ig; 1.

CC TIGRFAMs: TIGR01167; LPXTG_anchor; 1.

CC TIGRFAMs: TIGR01168; YSTRK_signal; 1.

CC PROSITE: PS50847; GRAM_POS_ANCHORING; 1.

CC Cell wall; Peptidoglycan-anchor; Receptor; Signal;

KW Immunoglobulin domain.

```
FT SIGNAL 1 37
FT CHAIN 38 1135
FT PROPEP 1136 1164
FT DOMAIN 434 534
FT DOMAIN 199 438
FT DOMAIN 439 526
FT DOMAIN 827 945
FT SITE 1132 1136
FT MOD_RES 1135 1135
FT SEQUENCE 1164 AA; 65DE94AF720A5474 CRC64;

Query Match
Best Local Similarity 5.1%; Score 141; DB 1; Length 1164;
Matches 126; Conservative 95; Mismatches 207; Indels 202; Gaps 32;

QY 2 KDLDKKEIKI-----AGIMK--QGVKRESIVVAKENAILIYPHGDHHDPI---48
DB 240 EDAEVKREELGKLFSSSTRAGLDQETOEHVKET---SSEENT---QKVDEHYANSLONL 293

QY 49 -----DEHKPVGIGHSHSNYELFKPEEGVAKKGNKYV--TGELTNVWVLL 93
DB 294 AOKSLEELDKATTNEQATQVKNQFLENACKLKEIQLIKETNVKLYKAMSESLEQVEXEL 353

QY 94 KNSTFNQNFTHANGKRVYSFPEPELEKLGILNMLVLTDPGKVLKVKSGKVFGEVGV 153
DB 354 KHNSEANLEDLVAKSEIVR-----EYEGKL--NQSKNL--PELKOLDEEAHSHKIQVVE 404

QY 154 NI-ANFELDQPYLPQOTFKYTIASKD-----YPE--VSYDG---TFTVPTSLEYK 197
DB 405 DFRKFKETSEQVTPKRVKVRDLAANENNQOKTELAVSPENITVYEGEDVKFTVTA---K 460

QY 198 MASQTIFFPHAGDTYLYRNPQFVAPKGTDALRVFDEPHGNAYLENKYKGEIKLPIPK 257
DB 461 SSKT---TLDFSLLTKYNPSVS-----DRISTNYKNT---DNHKTAEITIKWLK 506

QY 258 LNOQTTRT-----AGNKIPVTFMANAYLDNQSTVIVEVPILEKENQTDKPSILPQKRN 311
DB 507 LNESQVTLKAKDSDGNNVEKTF-----ITVQKKEE-----QVPKT 544

QY 312 KAQENSKLDEKV-EPPKTESEKVEKELSETGN-----STNSSTLEEVPTV-----355
DB 545 PEKDSKTEKYPQEPKSNKDKNOLOELKSAQOELKLEKAIKELMEQPEIPSNPEYCIQ 604

QY 356 -----DPVQEKVAKF-----AESYGNKLENVLFNMDGTIELYLPSEGVK 395
DB 605 KSIWESQKPEIQEATSFKKIIGDSSSKYTYEYHFNKYKSDPMNQLHAQM-----EMLT 659

QY 396 KKNADFTGAPQNGENKPSNGKYST-----GTVENQFTE-----NK-----433
DB 660 RRVQYVMNKPDPNAETKTFESDMKRTKEDNYGSLNDALKGYYFKYFLTPFNKIKQIVD 719

QY 434 PADSLPEAPNEKVPKPNSTNGMLNPEGNVSGDPMLOPALDEA-----PAVDPVQ 484
DB 720 DLDKVEQDQAPI-PENSE-----MDQAKKAKIAYSKYMSKVLGDGVH 762

QY 485 EKLEKFTASYGLGL-----DSVIFNMD 506
DB 763 QHLQKKNNSKIVDLFKLEDAIKQOITFDID 792

RESULT 5
SCA4_RICRH STANDARD; PRT; 1013 AA.
AC Q9AJ81;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia rhipicephali.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
```

```
OX NCBI_TaxID=33992;
RN [1]
RA SEQUENCE FROM N.A.
RA Sekevoya Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RL "gene D" coding for an intracytoplasmic protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF155053; AAK30684.1;
KW Antigen.
FT NON_TER 1013 1013
FT NON_TER 1013 AA; 110550 MW; 856E98912315D102 CRC64;
SQ SEQUENCE 1013 AA; 110550 MW; 856E98912315D102 CRC64;

Query Match
Best Local Similarity 5.1%; Score 140; DB 1; Length 1013;
Matches 113; Conservative 71; Mismatches 162; Indels 150; Gaps 26;

QY 48 IDEHKPVGIGHSHSNYELFKPEEGVAKKGNKYVGTGELTNVWVLLKNSTFNQNTLAN 107
DB 428 IDSNKQIDLLIKERATAILHNAKSDIAEKQ-----TNIITAEEN-TVNNQNL---473

QY 108 GQKRVYSFPEPELEKLGILNMLVKLI-----TPDGKVLKVKSGKVFGEVGNIANPELQ 162
DB 474 -----PD-AKAVGVNAVLETIKNDQNTPD---LER--SKMLEATVAITLSENLE 517

QY 163 PYLEPQOTFKYTIASKDYPEVSYDGTFTVPTSLEYKMASQTIFFPHAGDTYLYRNPQFV 222
DB 518 PKQKEOMLEK-----TVDVGLSLK-----DDASRA---AAI 545

QY 223 PKGTDALRVFDEPHGNAYLENKYKGEIKLPIPKLNOQTTRTA-GNKIPVTMANA---278
DB 546 DGIITDAV-----IKSNLSTEDKGTMLIAGDKVNVSELSNAEKQ 584

QY 279 -YLDNQSTVIVEVPILEKENQTDKPSILPQKRNKAQENSKLD--EKVEP-----325
DB 585 KLLGSVLKKGVEAQVLSPAQO---QLMQQNLDKITAEOQTKKOTIKKVNILFDPLSSTE 640

QY 326 -----PKTSEKVEKELSETGNSTNSTLE---EVPVDPVQEKVAKFA 366
DB 641 LKTTNIQAITSNVLDGPATAE-VKGEITIQITNTVAGSSLEAQDKAEIVKGVGETIATHS 699

QY 367 ESYGNKLENVLFNMDGTIELYLPSEGVKIKKNMADFTGEAPQNGENKPSNGKV 420
DB 700 DT-SLSLPKALIMASAEKGLIVESKTNLPBELMTKGLVDGIYE---GKGGPEITKAV 753

QY 421 STGTVENQPTENKPADSLPEAPNEK-----PVKPNSTNGM--LNPEGVGSDPMMDPA 473
DB 754 SSG-IDNSINDSEKALKRAKDAASEATLDIETQNLTE-CLKGQNTIEHKPRDDIYNKA 811

QY 474 LEEAPAVDPVQEKLEK 489
DB 812 QEVINAVNPVIEALEK 827

RESULT 6
TP2A_CRIGR STANDARD; PRT; 1526 AA.
AC P41515;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II, alpha isozyme (EC 5.99.1.3).
GN TOP2A OR TOP2 OR TOP-2.
```

Db	839	WYIPIPMVLNG	--AEGIGTWGCKIPNFDIRVWNIIRRLDGEELPMLSYAKFK	895
QY	172	YTI----	ASKDY--PEVSYDGTFTYPTS-LAYKMASQITFYPPFHAGDYLRL--VNPQFAV	222
Db	896	GTIEELASQNYINGEVA	LNSTTIEISLPTWTQ-----TYKEQVLEPMLNG	945
QY	223	PKGTDLVRVDFEHGNAY	LENNYKVGELK-----DIPKLNQGTTFAGNKIPVT	273
Db	946	TEXTPLLIIDYREYHND	TVKFIKMTTEKLAEAERVGLHKVFKLQ--TSLTCNSMWLFD	1003
QY	274	FMA-----	-----NAYLDNQSYTIVEVPILFK-	295
Db	1004	HVGLCKKYDVLIDLKOF	ELRIKYGLRKEWLLGMLGAESAKLNQARF-----ILEKI	1058
QY	296	-----	ENOTDKPSILPQFKRKAQENSKLDKVEEPTSEKVKEKULSETGNSSTL	349
Db	1059	DGRIITENPKKELIKVL	IQRGYDSDPKAWKAAQ-KVPDEEENE--ESDNENSDVA	1114
QY	350	EEVPTVD-----	PV-----QEKVAKFAESYGMKLE--NVLEVMNDDT-----IELVPSGE	392
Db	1115	ESGPTNYLLDNLPLW	LITLKEKDELCKQRNEKEQELTNKKNSPSDLWKEDLAVPIELE	1174
QY	393	VIK-KNMAFTGEAP	CNGENK-----PSNG-----KV	420
Db	1175	WBEAKBKDEQVGLPG	KGGKAKGKAQMSVLPSPHGKRVIPQVTMEMKAEKKIRKKI	1234
QY	421	STGTVENQPTEN-----	KPADSLPEAPNE-----KPVK-----PENSTDN	455
Db	1235	KSNVSGTPTENGELG	SLIKQRIEKKQKKEPQAMTKGTTIAFKPIKKGKKNRNPWSDES	1294
QY	456	GMLNPEGNVGS	DMLDPALAEAPVDPVQEKL-----EKFTASYGLGLDSVIFNMDDGT	508
Db	1295	DMSSNSNVDVPR-----	EKPRAATAKAKTMTDLDSDEPFGSDGKDDEDFPDLDTT	1349
QY	509	IELRPSGEVIKKN	522	
Db	1350	-----	PPKTKIPQKN	1359
RESULT 7				
TCN3_RAT	ID	TCN3_RAT	STANDARD;	PRT; 357 AA.
AC	PI9814;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Trans-Golgi network integral membrane protein TGN38 precursor.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
NCBI_Taxid	10116;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN-Sprague-Dawley; TISSUE=Liver;			
RC	MEDLINE=90372941; PubMed=2204342;			
RX	Luzio J.P., Brake B., Banting G., Howell K.B., Braghetta P.,			
RA	Stanley K.K.;			
RA	"Identification, sequencing and expression of an integral membrane			
RL	protein of the trans-Golgi network (TGN38).";			
RL	Biochem. J. 270:97-102(1990).			
RN	[2]			
RN	INTERACTION WITH NEURABIN-I AND NEURABIN-II, AND MUTAGENESIS.			
RP	MEDLINE=99445568; PubMed=1051494;			
RP	Stephens D.J., Banting G.;			
RT	"Direct interaction of the trans-Golgi network membrane protein,			
RT	TGN38, with the F-actin binding protein, neurabin.";			
RL	J. Biol. Chem. 274:30080-30086(1999).			
CC	-i- SUBUNIT; INTERACTS WITH NEURABIN-I AND NEURABIN-II. BINDS			
CC	PREFERENTIALLY TO THE DIMERIC FORM OF NEURABIN-I.			
CC	-i- SUBCELLULAR LOCATION: TRANS-GOLGI NETWORK.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collabora			


```

FT REPEAT      674      696      4-2.
FT CONFLICT    216      216      S -> P (IN REF. 2).
FT CONFLICT    491      491      E -> Y (IN REF. 2).
FT CONFLICT    514      514      L -> H (IN REF. 5).
SQ SEQUENCE    710 AA; 77856 MW; 90688ACAE6BDB334 CRC64;

Query Match
Best Local Similarity 4.8%; Score 133; DB 1; Length 710;
Matches 134; Conservative 70; Mismatches 246; Indels 232; Gaps 27;

QY 3 DLDKTEKIEAGIKOYGVKRSIVVNKEK-----NAILPHGCHHAPDIDHKPV---G 55
Db 64 DDDLEPEVIDAPGTGPRINTPASEEILPPGKTFPVVSSDYTKP-TESVPOQAS 122
QY 56 IGH---SHS-----NYELFKPEEGVAKKGNKVTGELTNVNLKNSTFN 99
Db 123 YGHDAFAHVRITFTSKDEEKRDVPIHHLSELSORESETHESLNTVPVLLSGTSDV 182
QY 100 NQNT-----LANGKRVSFEP-----PE 119
Db 183 TSTFAPSGDDEYLDGQKYNVETPIILEESAVSYLGSVNYQSKVTDPTKEETGGVPE 242
QY 120 LEKKLGINKLVKITPDGK-----VLEK-----VSGKV 147
Db 243 IAESFG-NNEVTDSEDPQGFQERDLSTRKEFEFDQDFDSVLCKDSPAKFPGESGV 301
QY 148 FEGVGNIANFELDQPLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQIFYPF 207
Db 302 FPGVGDSEGALE-----KDFPTRSHDFDMKTETGMDTNSPSRSHFEDL 346
QY 208 HAGDTYLRVNPQ-FAVPGKTDALVRVEDEFGHNAVLENNYKVGIEKLPKLNQGTTRTA 266
Db 347 KTESGNDKNSPMFGSESGAE-LEKFDKNDSG--RNEYS-----PESDGLGAPL 395
QY 267 GKNKIPVTMANAYLDNQSYTIVEVPI-----292
Db 396 GGNFPVR-SHELDLKNESDIDKVDVPGFGEPDFLAKGRPGYGEASEDKFPARSDDVEV 454
QY 293 ---LEKENTD-----RPSILPQKRNKAQENSKLDEKVEPKTSEKVK--EKL 337
Db 455 ETELGRDPKTEILDQSPSLSPKPERDEKESRDDFEETRDEKTEEPKOSTYTERFASML 514
QY 338 SETGNSNSTLSEVPTD---PVQEKVAKFAESYGMKL-----ENVLF 378
Db 515 GYSGEIPVGDQTVAGTVDEKLTPTVNEKDQETESAVTTKLPSGGSGGVFEORGEDKSVS 574
QY 379 NMDGTTILPDSGEVKKKNMADETFGAPCGNGENKPSNGKVGSTGTVENQPTENPADSL 438
Db 575 GRDYVAEKLTTTEE--DKAFSDMVAEKLQIGGEEKKE---TTTKEVEKISTE-KAASEE 628
QY 439 PPAPNEK-----PVKPN-----STDNGMLNPEGNGSDPML 470
Db 629 GEAVEEVKGGGMVGRKIGWFGGGATDVKPPSHSVSEAPKSSGWFGGGATEEVKPKS 688
QY 471 DPALREAP-----AVDPQEKL 487
Db 689 PHSVEESQSLGSTVVPVQKEL 710

RESULT 9
NFM_CHECK
ID NFM_CHECK STANDARD: PRT; 857 AA.
AC P16053;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M).
GN NFM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.

```

```

NCBI_TaxID=9031;
[1]
SEQUENCE FROM N.A.
MEDLINE=90174973; PubMed=2106668;
RA Zopf D., Dineva B., Betz H., Gundelfinger E.D., Betz H.;
RT "Isolation of the chicken middle-molecular weight neurofilament
(NF-M) gene and characterization of its promoter.";
RL Nucleic Acids Res. 18:521-529(1990).
[2]
SEQUENCE OF 259-857 FROM N.A.
MEDLINE=88112814; PubMed=3123320;
RA Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D., Betz H.;
RT "Identification of gene products expressed in the developing chick
visual system: characterization of a middle-molecular-weight
neurofilament cDNA.";
RL Genes Dev. 1:696-708(1987).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC -!- PM: WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; X17102; CAA34958.1; -
EMBL; X05558; CAA29073.1; -
PIR; A27040; A27040.
PIR; S08061; S08061.
PIR; S15762; S15762.
InterPro; IPR001664; IF.
Pfam; PF00038; Filament; 1.
PROSITE; PS00226; IF; 1.
Intermediate filament; Coiled coil; Neurone; Phosphorylation;
Glycoprotein.
INIT_MET 0 0
DOMAIN 1 98 HEAD.
DOMAIN 99 406 ROD.
DOMAIN 407 857 TAIL.
DOMAIN 99 130 COIL 1A.
DOMAIN 131 143 LINKER 1.
DOMAIN 144 242 COIL 1B.
DOMAIN 243 259 LINKER 12.
DOMAIN 260 281 COIL 2A.
DOMAIN 282 285 LINKER 2.
DOMAIN 286 406 COIL 2B.
CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
CARBOHYD 426 426 O-LINKED (GLCNAC) (BY SIMILARITY).
CONFLICT 546 546 G -> R (IN REF. 2).
SEQUENCE 857 AA; 95704 MW; 4E2E0FC6AC64778B CRC64;

Query Match
Best Local Similarity 4.8%; Score 132.5; DB 1; Length 857;
Matches 62; Conservative 29; Mismatches 121; Indels 43; Gaps 7;

QY 249 GEIKLPIPKLNGQTRTAGNKIPVTFMANAYLDNQSYTIVEV-----PILEKNTDKPS 303
Db 411 GSITGPITHTQPSVTIASTKIQTKEIPPKLVQHAFVEIIEETKVEDEKSEMEDALS 470
QY 304 ILPQPKRNKAQENSKLDEKVEEPTSEKVEKLSKTSNLTSETSNLTLEVPVDPVQEKVA 363

```

```
Db 471 ALAESNAKAEQEEBEEKAEAEVEEBAVSEKRAEQAAEEBEEBEEAE-----EEBAA 525
Qy 364 K--FRESYGMKLVNLFNMDGTIELYLSGEVVKKNMADFTGEAPQNGENKPSNGKVS 421
Db 526 KSDAAEEGSKKEEIEKEEGS-----EAEDEEAEAK-----GRAE 561
Qy 422 TGTVENOPTENKPADSLPEAPNEKPKVKNSTNGMLNPEGNVGSDPMDLPALBEA---- 477
Db 562 EAGKVEKVKSPKSPKSPKSPKSPVTEQ---AKAVQKAAAEVKGDKQAKAEKAAKEE 618
Qy 478 PAVDPVOEKLKFTTA 492
Db 619 KAASPEKPATPKVTS 633

RESULT 10
MSPL_PLAFM STANDARD; PRT: 1701 AA.
AC P08569;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=70153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum.";
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP REVISIONS TO 1403: 1569 AND 1629.
RA Tanabe K.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
CC -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).
CC -|- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X05624; CAA29112.1;
CC PIR; A26868; A26868.
CC PIR; B25120; B25120.
CC InterPro: IPR000561; EGF-like.
CC Pfam: PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
GN SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 4.8%; Score 131.5; DB 1; Length 1701;
Best Local Similarity. 19.8%; Pred. No. 8.8; Mismatches 258; Indels 105; Gaps 20;
Matches 112; Conservative 95;

Qy 3 DLQKKEKIA-----GIMKQYGVKRESIVVNVKKNIAI---YPHGDHHADPI 48
Db 435 DTEKINEKIITDNKERKIFINNIKKQIDLEKNINHTKEQNKLLDEYKSKDYELL 494
Qy 49 DEHPVGIHSHSNYELFKPEEGVAKKEGNKVTGEBELTNVWLLKNSTFNNOFTLANG 108
Db 495 EK-----YEM-KFNNDKDVQKIFESARTYVYKORYNNKFNSSNSVYV 542
Qy 109 Q---KRVSEFPPELEK-----KGINMLVKLITPDGVLE-KVSGKVFEG 151
Db 543 QKLKALSYLEDYSLRKGISEKDFNHYYTLKTGLEADIKLKEIKSENKILEKVF-KG 601
Qy 152 VGNIANFELQPYLPQGTFKYTIASKDYPEVSYDGTETVPTSLAYKMASQTIYFPHAGD 211
Db 602 LTHSANASLEYSDIVKLQVQVLLIKKIEDLRKTELFLKNAQLKDSIHVPNIYKPNKPE 661
Qy 212 TYLRVNFQFVAPKTDALVRFDEFHONAYLNNYKYEIKLPIPKLNQGTTRTAGNKIP 271
Db 662 PYLIVLKKEVDKLEKFIKPKVKMLK-----KEQAVLSSITQPLVAASE-TTEDGGHSTH 715
Qy 272 V-----TFMANAYLDNOSTYIVEVPILEKENQTDKPSILPQFKRNKAQENSK- 318
Db 716 TLSQSGETEVTETETVETVGHHTTITLPPKE-ESAPKEVKVVENSTEHKSNDNSQA 774
Qy 319 LDEKVEBKTSKVEK-----EKLSETGNSISNLEVPVTPVQEVKAFASYSKMKE 374
Db 775 LTKVYLKLDLFTKSYCHKYILVSNSSMDQKLLLEVYLTPEEEKELKSCDPLD-LF 833
Qy 375 NVLFNMDGTIELYLPQG-----EVIKKNMADFTGEAPQNGENKPSNGKVSQTGIVE 426
Db 834 NQNNIPAWSYLSDMNDLQHLFFELYQKEMIVYLKLEENHIKLLBEQKQITGSS 893
Qy 427 NQPTENKPADSLPEA-----PNEKPVKPENSTDNGMLNPEGNVGSDPMDLPALBEAFAVD 481
Db 894 TSSPGNTVTNTAQATHSNQSQNSASSTNTQNGV-----AVSSGPAV- 937
Qy 482 PVOEKLEKFTASYGLGLDSDVIFNMDGTIEL 511
Db 938 -VEESHDPDLTVL-----SISNDLKGVSL 960

RESULT 11
GLYB_PLAFG STANDARD; PRT: 774 AA.
AC P02895;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glycophorin-binding protein precursor (GBP-130).
GN GBP.
OS Plasmodium falciparum (isolate FCR-3 / Gambia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5838;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86133561; PubMed=3512098;
RA Kochan J., Perkins M., Ravetch J.V.;
```

```

Db      158 NOKEGKNVKVIRKKEKKESGGPEENKHANEASKKPRASKYSQKPSTSTRGNNEVKI 217
           || : :: || : | : | | | : || : | : | : || : | : | : || : | : | :
QY      344 TSNSTLEEVPDYPQVKAKFA--EYGMKLE---NVLFNMDGTIELYLPSGEVIKK-- 396
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      218 RAASNQETLISADPEGQIMREYAADPYRKHLEFYKLINTD-----PNDEVERNA 270
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      397 -NNADFTGEAPQGNGENKPKSPENG-----KVSTCTVENQPTEKNPADSLPAPNEK 445
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      271 DNKEDLTSDAPEQIMREYASDPYRKHLEIFYKLINTDNDDVERRNAD----- 321
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      446 PVRPENSTDGMLNPENGVSDPMLDLAEAPAVDPVOEKLEFNTASYGLGLDSVFNM 505
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      322 -----NKEDTSDAPEQIMREYAADP-----EYRKHLEVFI-----HKILTT 359
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      506 DGIELRLPSGEVIKKN 522
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      360 D-----PNDEVERRN 369
           || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :


RESULT 12
CNA_STAAU STANDARD; PRT; 1183 AA.
ID CNA_STAAU ID AC Q53654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=921165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RT "Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin.";
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2] ERRATUM.
RN Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3] COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boies J.O., Hoeoek M.;
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus.";
RL Biochemistry 32:11428-11435(1993).
RN [4] X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF I69-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., DeLucas L.J., Hoeoek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin.";
RL Nat. Struct. Biol. 4:833-838(1997).
CC -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC purposes without permission is prohibited.

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M81736; AAA20874.1;
CC PDB: 1AMX; 24-JUN-98.
CC INTERPRO: IPR001899; Gram_pos_anchor.
CC TRIGRAMS: TRIGRAM1167; LPXTG_anchor; 1.
CC PROSITE: PS0847; GRAM_POS_ANCHORING; FALSE_NEG.
CC Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
CC SIGNAL 1 29
CC CHAIN 30 1154
CC PROPEP 1155 1183
CC DOMAIN 151 318
CC DOMAIN 533 1093
CC DOMAIN 1093 1157
CC REPEAT 533 719
CC REPEAT 720 906
CC REPEAT 907 1093
CC SITE 1151 1155
CC MOD_RES 1154 1154
CC SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

Query Match 4.8%; Score 130.5; DB 1; Length 1183;
Best Local Similarity 21.6%; Pred. No. 6;
Matches 99; Conservative 62; Mismatches 197; Indels 101; Gaps 19;

QY 74 KKEGKV-VTGEELTNVLLKNSTFNQFTLANGQKRVSFSPPELEKLGINLVKL 132
DB 680 KARGQVKYVEELTKV-----KGYTHVDNNDM--GNLIVTNKTYPTETTSIGERKVVDDK 733
QY 133 ITPDGKVKLEKSVKVEG-----VGNIANFELDQVLP---GQTFKYTIAS---KD 178
DB 734 DNQGRPEKSVNLLADGEKVKTLDTSETNKKYEFKDLPRYDEGKKTEYIVTVDHVKD 793
QY 179 YPEVSDGTFTVPTSLAYKMSQTIYPFHAGDTLVY-----NPQFAVPKGTDAIVRV 232
DB 794 Y-TTDINGT-----TIINKYTPGETSATVTKNNDNNQDGKRPTEIKVEL 838
QY 233 FDE-----FHGNAYLENNKYGEIKLPIPKLQ---GTRTAGNKKIPVT 273
DB 839 YQDKATGKATAILNESNNTHWTGLDEKAKGQOVKYTYVEELTKVGYTHVDNMDGNL 898
QY 274 FMANAYLDNOSTVIVPEILEKENQDK-----PSILPQFRNKA-----Q 314
DB 899 IVTNKYTPETTSISGEKVDKDNQDGKRPKSVNLLANGKVKTLDTSETNKKYEFK 958
QY 315 ENSKLDE--KVEPKTSKVEKEKLSCTENSTNSNLT--EEVPTVD-----P 357
DB 959 DLPKYDEGKKIETVTDHVKDYTTDQINGTITINKYTPGETSATVTKNNDNNQDGKRP 1018
QY 358 VQEKVAKFAESYGMKLENVLNFMNDGTIELYLPFSGEYIKKNMADFT--GEAPQNGENKPS 416
DB 1019 TEIKVELYQDKATGKATAILNESNNTHWTGLDEKAKGQOVKYTYVEELTKVNGYTHVD 1078
QY 417 NGKVSCTGVENQTEKNKPADSL-PEAPNEK--PVKPENS 452
DB 1079 NNDMGNLIVTNKYTPKKPNKPIYPERKPKDKTPTPRDHS 1117

RESULT 13
Y832_METJA STANDARD; PRT; 1750 AA.
ID Y832_METJA
AC Q58242;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0832 [Contains: Mja rnr-1
DE intein].
DE MJ0832.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaeae; Methanocaldococci.
OX NCBI_TaxID=2190;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervilave A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uutterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- PFM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC -!- SIMILARITY: WEAK IN THE C-TERMINAL, TO M.JANNASCHII MJ0885.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67527; AAB98834.1;
CC HSP: P17255; IDFA.
CC TIGR: MJ0832;
CC InterPro: IPR005144; ATP.
CC InterPro: IPR003586; Hedgehog_hintC.
CC InterPro: IPR003587; Hedgehog_hintN.
CC InterPro: IPR002203; Intein.
CC InterPro: IPR004042; Intein_endonuc.
CC Pfam: PF03477; ATP_cone; 2.
CC PRINTS: PR00379; INTEIN.
CC SMART: SM00305; HincC; 2.
CC SMART: SM00306; HincN; 2.
CC PROSITE: PS50818; INTEIN_CTER; 2.
CC PROSITE: PS50819; INTEIN_ENDONUCLEASE; 2.
CC PROSITE: PS50817; INTEIN_NTER; 2.
CC Hypothetical protein; Autocatalytic cleavage; Protein splicing;
CC Complete proteome.
CC CHAIN 1 337
CC CHAIN 338 790
CC CHAIN 791 1058
CC CHAIN 1059 1591
CC CHAIN 1592 1750
CC SEQUENCE 1750 AA; 203279 MW; 6D68938770A25359 CRC64;

Query Match 4.8%; Score 130.5; DB 1; Length 1750;
Best Local Similarity 20.2%; Pred. No. 10;
Matches 124; Conservative 113; Mismatches 225; Indels 151; Gaps 36;

QY 1 MKDLDK----KIEEKIAGIMKQYGVKRESIVYVNEKNAIYPHGDHHDHAPIDHEKPVGI 56
DB 347 IKEGSKLVKVGKIGEAIDFEMEKY---KDKIIVDGDTEIL-----YLDGIAEVYIVSV 395
QY 57 GSHSNYELFPKEECVAK-KEGNKVY--TGELTNVLLKNSTFNQFTLANGQKRV 112
DB 396 NVKTKGAK-PRVYVAISRKHPRGVKGVIGKDGTSIIIVTDEHSLFNVDENGLNVCVPRQ 454
QY 113 -----SFSFPELEKLGINLVKLITPDCK-----VLEKSVKVFGEVGNIAF 159.
DB 455 MKHIIRNENFYDVEYRIGDYIETNYQFTDSKYNQRNDIPEKL--KITKELCQFLGV 512
QY 160 LDQVLPQGTQKTYTASDKYPEVSDGTFT---VPTSLAYKMSQTI-----FYFP--- 207
DB 160 LDQVLPQGTQKTYTASDKYPEVSDGTFT---VPTSLAYKMSQTI-----FYFP--- 207

```



```
Db 513 AGSVI---TNGISITTKDDDDIAKIEFVKEQINENAVKRYEDSVRFVNGFYRFLKE 569
Qy 208 HAGDTYLRVN-PQFAVPKGTDALVRV--DEFHGNAYLENNYKV-----GEIKLP 254
Db 570 HINGRAINKNSPEF-ILKGDREKMLAFLGLGISDGYVSKDGRVQIYVTSQQLGLQLHL 628
Qy 255 IFKLNQ----GTRTRAGNKIPVFMANAYLDNOSTYIVVEVILEKENOTD--KPSILPQF 308
Db 629 LSDLOMIVSITKIEEGKIEF--KRNEIVRNKYLVIEI-----AKNCTEDLKYVIEKY 682
Qy 309 KRNKAQ----ENSKLDERVEEPEKTEKVEKELSETGSTGN-----STLBEVPTVDV-V 358
Db 683 KKRERKPNYDQLPYDRIKIEHLKRIITDKKPYNDYANKSNRKLKLTLEKIEQLNPHL 742
Qy 359 QKVKAFBESYGMKLENVLFNMDGTIELYLPDSGEVVKKNMADFTGEAPQGNENKPSNG 418
Db 743 REEINKF-----KL-NIPFEIKEKEI-----DYGIVYVDLSVED--NENE 780
Qy 419 KYSTGTVENQPTENKPADS-----LPAPNEKPKVKPENSTDNM-----457
Db 781 ITATGIL-----CHNIFISINLEIPELFDKPAVAGTTRGTYGDYEEAKLILEALV 836
Qy 458 -INPGNVGSDPMLDP-----ALDEAPADVPQVE---KLEKFTASVGLDLSVFNMDGTI 509
Db 837 DVMMEGDAMGKFFLEPNFTIKLRNFAFKDENKELMYKIHQLSAKEGI---PYFINN----889
Qy 510 ELRLPSGEVIKKN 522
Db 890 ---LPDWQVNTN 899

RESULT 14
ID IGA2_HAEIN STANDARD; PRT; 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / Serotype B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RL influenzae type 1 immunoglobulin A1 proteases.";
CC J. Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
```

```
DR EMBL; M87489; AAA24966.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Perfact_sup.
DR Pfam; PF02395; IGA1; 1.
DR Pfam; PF03212; Perfactin; 2.
DR PRINTS; PRO0921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Repeat; Signal;
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
FT REPEAT 1109 1116 K.
FT REPEAT 1117 1124 2.
SQ SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

Query Match 4.7%; Score 130; DB 1; Length 1702;
Best Local Similarity 20.5%; Pred No. 11;
Matches 106; Conservative 61; Mismatches 201; Indels 150; Gaps 20;

Qy 72 VAKKEGKVV-----YTGEELTNVYNLLKSTFNQNTFLANGOKRVSFSPPELEKLG 126
Db 783 IGYKAGDTVCVRSDYTG-YVTCTTDLKSLDKALNSFNATVSGNVNLS-----GNANFVLGK 837
Qy 127 NMLYKLITPDG----KVLKVSQKVGCEGVGNIAFELDQPYLPQGTFKYTIASKD---Y 179
Db 838 ANLFGTISGTGNSQVRUTENSHWHLTGDSNVNQLNKGHIHLNAQDANKVITYNTLT 897
Qy 180 PEVSYDGTFTVPTSLAYKMASQTTFFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEPHGN 239
Db 898 NLSGNGSFYLLDLSNKGQDKVVTGATGNFTLQVADKTGEP--TKNELTLFDA--SN 953
Qy 240 AYLENNYKVGKILPIPKLNOGTTTRTAGNKLPVTFMANAYLDNQSTYIVVEVILEKENQT 299
Db 954 A-TRNNLNVSLV-----GNTVDLGAMKYKLRNVNRYDLYNPEVEKRNQT 997
Qy 300 -----DKPSI-----LPQFKRNKAQNSKLDER---322
Db 998 VDTNITPNNIQADVPSVPSNNEEIARVETPVPPAPATPSETTEITVAENSQESATVE 1057
Qy 323 -----VEEPKTSK--VEKEKLSKETGNSNSTLEEVPTVPVQ--EKVA 363
Db 1058 KNEQDATETTAQNGEVAEEAKPSVKANTQTNEVAQSGSETEETQTTEIKETAKVEKEKA 1117
Qy 364 KFAESYGMKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPQGNENKP-----414
Db 1118 KVEKEKAKVEK-----DEIOEPQMASSETSPKQAKPAPKE 1153
Qy 415 -SENGKYSTGTVENQP-----TENKPADSL--PEAPNEKPKYPK---ENSTD 454
Db 1154 VSTDKVEETQVQAQPQTQSTTVAABEATSPNSKPAEETQPSKTNAPETPVTVVSKNQTE 1213
Qy 455 NGMLNPEGNVGSDPMLDPALEEAFAV---DPVQEKLE 488
Db 1214 NTTDQPTEREKTAKEVTEKTEQPPQVQASQAPKQEQSE 1251

RESULT 15
YNR3_YEAST
ID YNR3_YEAST STANDARD; PRT; 366 AA.
AC P3885;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 40.3 kDa protein in RPS3-PSD1 intergenic region.
GN YNL173C OR N1673.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
```


OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:07 ; Search time 23.447 Seconds
(without alignments)
4639.948 Million cell updates/sec

Title: US-09-471-255-10
Perfect score: 2746
Sequence: 1 MKDLKKTEETAGIMKQYG.....TELRLPGEVKKNLSDPIA 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries.

Database :

- SPTREMBL_21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvrius.*
 - 16: sp_bacteria.*
 - 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2734	99.6	1039	16 Q9ANY1	Q9any1 streptococc
2	184	6.7	2004	16 Q37QP7	Q97qp7 streptococc
3	169.5	6.2	1964	2 Q59947	Q59947 streptococc
4	155	5.6	1524	10 Q8RYN2	Q8ryn2 oryza sativ
5	153	5.6	1873	2 Q924N7	Q924n7 enterococcu
6	149.5	5.4	4888	16 Q8PQ08	Q8pq08 ureaplasma
7	148	5.4	1939	5 Q25662	Q25662 plasmodium
8	147	5.4	2151	5 Q9NG79	Q9ng79 trichomonas
9	146.5	5.3	791	3 Q9USW1	Q9usw1 schizosacch
10	146	5.3	1272	16 Q98PR8	Q98pr8 mycoplasma
11	145.5	5.3	1078	5 Q963T1	Q963t1 plasmodium
12	145.5	5.3	1927	2 Q54875	Q54875 streptococc
13	142	5.2	2485	5 Q96134	Q96134 plasmodium
14	141.5	5.2	1441	16 Q9CFL1	Q9cfl1 lactococcus
15	141	5.1	1134	2 Q99051	Q99051 streptococc
16	140.5	5.1	17352	5 Q95YM2	Q95ym2 procambarus

17	140	5.1	841	3 Q94002	Q94002 candida alb
18	139.5	5.1	2443	5 Q9VSA2	Q9vsa2 drosophila
19	139	5.1	750	16 Q9PP44	Q9pp44 campylobact
20	138.5	5.0	1127	5 Q94248	Q94248 caenorhabdi
21	137.5	5.0	387	16 Q97KL1	Q97kl1 clostridium
22	137.5	5.0	1377	3 Q42695	Q42695 candida alb
23	137.5	5.0	1526	11 Q55079	Q55079 crisetulus
24	137.5	5.0	2768	5 Q9VC00	Q9vc00 drosophila
25	137	5.0	940	16 Q928N7	Q928n7 listeria in
26	136.5	5.0	940	16 Q8Y4N9	Q8y4n9 listeria mo
27	135.5	4.9	881	2 Q9RFJ4	Q9rfj4 streptococc
28	135	4.9	825	2 Q93GT5	Q93gt5 streptococc
29	135	4.9	1255	2 Q9R2J6	Q9r2j6 shigella so
30	134.5	4.9	1049	5 Q9VI49	Q9vt49 drosophila
31	134.5	4.9	1764	2 Q93T34	Q93t34 haemophilus
32	134.5	4.9	2025	11 Q99PP2	Q99pp2 mus musculu
33	134.5	4.9	2481	16 Q99QR6	Q99qr6 staphylococ
34	134	4.9	380	11 Q63575	Q63575 rattus norv
35	134	4.9	2276	2 Q93TV6	Q93tv6 staphylococ
36	133	4.8	710	10 Q39060	Q39060 arabidopsis
37	133	4.8	1020	5 Q27184	Q27184 trichomonas
38	133	4.8	5458	5 Q9U459	Q9u459 plasmodium
39	132.5	4.8	806	16 Q9CKX5	Q9ckx5 pasteurella
40	132.5	4.8	1037	13 Q73808	Q73808 fugu rubrip
41	132.5	4.8	1526	11 Q55078	Q55078 crisetulus
42	132	4.8	1231	2 P72362	P72362 streptococc
43	132	4.8	1653	2 Q9L8G3	Q9lbg3 streptococc
44	132	4.8	2478	2 Q9LCH2	Q9lch2 staphylococ
45	132	4.8	2478	2 Q9RL69	Q9rl69 staphylococ

ALIGNMENTS

RESULT 1

Q9ANY1	PRELIMINARY;	PRT: 1039 AA.
ID Q9ANY1		
AC Q9ANY1		
DT 01-JUN-2001	(Tremblrel. 17, Created)	
DT 01-JUN-2001	(Tremblrel. 17, Last sequence update)	
DT 01-OCT-2001	(Tremblrel. 18, Last annotation update)	
DE	pneumococcal histidine triad protein E precursor (Hypothetical protein	
DE	SP1004).	
GN	PTE OR SP1004.	
OS	Streptococcus pneumoniae.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;	
OC	Streptococcaceae; Streptococcus.	
OX	NCBI_TaxID=1313;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21101045; Pubmed=11159990;	
RA	Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,	
RA	Dornitz M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,	
RA	Langermann S., Koenig S., Johnson S.	
RT	"Identification and characterization of a novel family of pneumococcal	
RL	proteins [the Pht family] that are protective against sepsis."	
RL	Infect. Immun. 69:949-958(2001).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RX	STRAIN=TIGR4;	
RA	MEDLINE=21357209; Pubmed=11463916;	
RA	Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,	
RA	Peterson S., Heidelberg J., DeBoy R.F., Haft D.H., Dodson R.J.,	
RA	Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,	
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,	
RA	Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,	
RA	McDonald L.A., Feldblyum T.V., Angiuoli S., Dickson T., Hickey E.K.,	
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,	
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;	
RT	"Complete genome sequence of a virulent isolate of Streptococcus	
RL	pneumoniae."	
RL	Science 293:498-506(2001).	
DR	EMBL; AF318956; AAK06761.1;	

```
DR EMBL: AE007403; AAK75121.1; -
DR TIGR: SP1004; -
KW Signal; Hypothetical protein; Complete proteome.
FT SIGNAL 29 POTENTIAL
SQ SEQUENCE 1039 AA; 114631 MW; 81A563FC806625C4 CRC64;

Query Match          99.6%; Score 2734; DB 16; Length 1039;
Best Local Similarity 99.6%; Pred. No. 1.3e-135;
Matches 526; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKDLDRKIEKIAIGKQYGVKRESIVVAKENAIYIPGHGHHADPIDEHKPVGIGHSH 60
Db 512 MKDLDRKIEKIAIGKQYGVKRESIVVAKENAIYIPGHGHHADPIDEHKPVGIGHSH 571

QY 61 SNYELFKPEEGVAKKEGKNKYITGEEITNVNLLKNTFNQNTFLANGOKRVSFSEPPPEL 120
Db 572 SNYELFKPEEGVAKKEGKNKYITGEEITNVNLLKNTFNQNTFLANGOKRVSFSEPPPEL 631

QY 121 EKKLGINMLVKLITPDGKVLKESYGVFGEVGNIANFELDPYLPQGTFFKYTIASKDYP 180
Db 632 EKKLGINMLVKLITPDGKVLKESYGVFGEVGNIANFELDPYLPQGTFFKYTIASKDYP 691

QY 181 EYSVGGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 240
Db 692 EYSVGGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 751

QY 241 YLENNYKVGEEKIPIKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIVVEVPILEKENQTD 300
Db 752 YLENNYKVGEEKIPIKLNQGTTRTAGNKIPVTFMANAYLDNOSTYIVVEVPILEKENQTD 811

QY 301 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSFGTNGSTNSTLEEVTVPDVOE 360
Db 812 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSFGTNGSTNSTLEEVTVPDVOE 871

QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEVVIKKNMADFTCEAPOGNGENKPSGNGV 420
Db 872 KVAKFAESYGMKLENVLFNMDGTIELYLPSEVVIKKNMADFTCEAPOGNGENKPSGNGV 931

QY 421 STGTVENQPTENKPADSLPEAPNEKPKYKPNSTDNGLNPNPEGVSDPMLDPALEAPAV 480
Db 932 STGTVENQPTENKPADSLPEAPNEKPKYKPNSTDNGLNPNPEGVSDPMLDPALEAPAV 991

QY 481 DPVQEKLEKTASYGLGLSDSVIFNMMDGTIELRLPSGEVVIKKNLSDFIA 528
Db 992 DPVQEKLEKTASYGLGLSDSVIFNMMDGTIELRLPSGEVVIKKNLSDFIA 1039

RESULT 2
Q97QP7 PRELIMINARY; PRT; 2004 AA.
ID Q97QP7
AC Q97QP7;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Immunoglobulin A1 protease.
GN SP1154.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TTG84;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickson T., Hickey E.K.,
RA Holt J.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
```

```
pneumoniae."
RT Science 293:498-506(2001).
DR EMBL: AE007416; AAK75263.1; -
DR MEROPS: M26.001; -
DR TIGR: SP1154; -
DR InterPro: IPR001899; Gram_pos_anchor
DR InterPro: IPR001130; Zn_MTpeptidse
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs: TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs: TIGR01168; TSIRK_signal; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease; Complete proteome.
SQ SEQUENCE 2004 AA; 223908 MW; 556BC6A1028D60A4 CRC64;

Query Match          6.7%; Score 184; DB 16; Length 2004;
Best Local Similarity 23.5%; Pred. No. 0.27;
Matches 130; Conservative 70; Mismatches 162; Indels 192; Gaps 34;

QY 60 HSNYELFKPEEGVAKKEGKNKYITGEEITNVNLLKNTFNQNTFLANGOKRVSFSEPPPEL 102
Db 279 HKNLET-KKEEKISPEKTVNTLNQDEVLGSLQINKPELLYREETMETKIDFOEEIQEN 337

QY 103 FTLANGOKRVSFSEPPPELKKLGINM-LVKLIT-----PDGKVLKES 144
Db 338 PDLAEGTVRV-----KOEGLKGLKKEIVRIFSVNKEVSEIYSTTAPSPRIVEKGT 391

QY 145 GK--VFGE---GV-----GNIANFELDPYLPQGTFFKYTIASKDYPEVSDGTFTVP 191
Db 392 KKTQVIKQPETGVGHKDVQSGAIVEPAL-QPELP---EAVVSDKGEPEVQ-----TLP 442

QY 192 TSLAYKMASQTIYFPFHAGDTYLRVNPQ---FAVPKGTDALVRVDFEHGNAYLENNYK 247
Db 443 EAVV-----TDKGET--EVQEPSPDTVSDKGEPEQVAPLPYKGN-----481

QY 248 VGEIK-LPIPKL-NOGTRTAGNKIPVTFMANAYLDNOSTYIVVEVPILEKENQTDKPSI 304
Db 482 IEQVAPETPEVETKQGPET--EEVPV-----KPTETPVNNEGTEGTSI 527

QY 305 LPQFRN---KAOENSKLDEKVEEPTSEK-----VEKEKLSFGTNGS 343
Db 528 --QEAENPVQPAEESTNTSEKV-SPDTSKNTGEVSSNPSTTSVSGSNKPEHNSKNE 584

QY 344 TSNSTLEEVTVPDVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVVIKKNMADFTG 403
Db 585 NSEKTVEEVP-VNP-----NEGTVE-----GTSNQET 610

QY 404 EAPQNGENKPSGNGV---STGTVENQPTENKP--ADSLPEAPNEKPKYKPNSE---TD 454
Db 611 EKPVPQPAEETQTNQSGKIANENTGVSNKPSDSKPPVEESNQPEKNGATKPNSENGTISE 670

QY 455 NGMLAPE---GNVGDSPMLDPALEAPAVDPVQEKLEKFTASYGLGLSDSVIFNMMDGTIEL 511
Db 671 NGQTPSPNGSTNDVSTESNTSNGNEEKQENE-----LDPKKKVEEPEKTELEL 723

QY 512 RLPSGEVVIKKNLS 525
Db 724 R-----NVSD 728

RESULT 3
Q59947 PRELIMINARY; PRT; 1964 AA.
ID Q59947
AC Q59947;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Immunoglobulin A1 protease (EC 3.4.24.13).
GN IGA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=R6;
RC  MEDLINE=97047667; PubMed=8926056;
RA  Wani J.H., Gilbert J., Plaut A., Weiser J.N.;
RT  "Identification, cloning and sequencing of the Immunoglobulin A1
RL  protease gene of Streptococcus pneumoniae.";
RL  Infect. Immun. 64:3967-3974(1996).
DR  EMBL: U47687; AAC44568.1; -
DR  MEROFS; M26.001; -
DR  InterPro: IPR001899; Gram_pos_anchor.
DR  InterPro: IPR00130; Zn_Mtpeptidse.
DR  Pfam: PF00746; Gram_pos_anchor.1.
DR  TIGRfams: TIGR01167; LPXTG_anchor.1.
DR  TIGRfams: TIGR01168; YSRK_signal.1.
DR  PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR  PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW  Hydrolase; Protease.
SQ  SEQUENCE 1964 AA; 218786 MW; 0599E818A783B7E8 CRC64;

Query Match          6.2%; Score 169.5; DB 2; Length 1964;
Best Local Similarity 23.7%; Pred. No. 1.5;
Matches 115; Conservative 60; Mismatches 140; Indels 171; Gaps 30;

QY  60 HSNYLFKPEEGVAKKEGKQVYI-----GEELTNVYN-----LLKNSTFN-----NQN 102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  279 HKNLET-KKEEXISPKEKTVGNILNPQDEVLSGQINKPELLYREETIETKIDFQFEIGEN 337

QY  103 FTLANGOKRVVSFPPELEKKGGINM-LVKLIT-----PDGKVLKVS 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  338 PDLAGTVRV-----KQGGKLGKVEIVRIFSVNKEEVSREIVSTSTTAPSPRIVEKGT 391

QY  145 GK--VFGE-----GV-----GNIANFELDQPLPGQTFKYTIASKDYPEVSYDGTFTVP 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  392 KKTQVIKEQETGVGHEKDVQSGAIVEPAI-QPELP-----EAVVSDKGPEVQP-----TLP 442

QY  192 TSLAYKMASQTFYPPHAGDTYLRNQP-----FAVPKGTDALVRVDFDFHGNAYLENYK 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  443 EAVV-----TDKGET--EYQPSPTVWSDKGEQVAPLPYKGN-----481

QY  248 VGEIK--LPIPKL-NOGTRTAGNKIPYTFMANAYLDNQSTIVIVPILKXENOTDKPSI 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  482 IQQVRPEPVEKTKQGGPEKT--EEVPV-----RPTETPNVNEGTEGTSI 527

QY  305 LPQFNRN---KQENSKLDEKVEEPEKTEKV-----EKKLSETGNST 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  528 --QEAENPVQPAEESTTNSEKVSPTDTSSENTGEVSSNPDSSTTSVGSNKPKPHNDSKNEN 585

QY  345 SNSTLEEYPTDVPQVEKVAKPAESYGMKLENVFNMDGTIELYLPSSGEVIKKNADFTGE 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  586 SKTVEEVP-VNP-----NEGIVE-----GTSNQETE 611

QY  405 APQGGNKPKSENGKV---STGTVENOPTENKP--ADSLPEAPNEKPKVPKPS-----TDN 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  612 KPQPAEETQTSNGRIANENTGEVSNKPSDSKPPVEESNQPEKNGTKATKPSNGNTISEN 671

QY  456 GMLNPE 461
Db  672 GOTEPE 677

RESULT 4
QRYN2
ID  Q8RYN2 PRELIMINARY; PRT; 1524 AA.
AC  Q8RYN2;
DT  01-JUN-2002 (TrEMBLrel. 21, Created)
DT  01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE  P0663E10.13 protein.
GN  P0663E10.13
OS  Oryza sativa (japonica cultivar-group).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
   OC  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

```

OC  Ehrhartoideae; Oryzeae; Oryza.
OX  NCBI_TaxID=39947;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CV_NIPPONBARE;
RA  Sasaki T., Matsumoto T., Yamamoto K.;
RT  "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RL  clone:P0663E10.";
RL  Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AP004317; BAB90730.1; -
SQ  SEQUENCE 1524 AA; 168701 MW; E47FE1A9C5842228 CRC64;

Query Match          5.6%; Score 155; DB 10; Length 1524;
Best Local Similarity 19.5%; Pred. No. 6.2;
Matches 107; Conservative 77; Mismatches 206; Indels 160; Gaps 22;

QY  9 EKIA--GIMKOYGVKRESIVNKR--EKNATYIPGHGHHADPIDEHKPKVIGHSHSN-- 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  498 DEKLTMRIRVKEETYEETALVNKNGDSSAVNLTHCDNEKGVLPQPKPQHTAQSGSKLE 557

QY  63 -----YELFKPEEGVAKKEGKQVYI-----GEELTNVYNLLKNSTFNNONFTLANGOKR 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  558 QLGKWTSGAEFYVLISPDQ---KCKTNSV-----TCFGDNVQTTNPSSK 598

QY  112 VSFSPPELEKKGGINMLKITPDGKVLKVSQKVFGEV---GNIANFELDQPLPG 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  599 LG-----QFEKKGK-----ETTSDFVGCCKSWDGGDIAELRHEHVNL- 637

QY  168 QTFKYTIASKD-----YPEVSYDGTFTVPTSLAYKMASQTFYPPHAGDTYLRVNPQF 220
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  638 ---EYAGTSTEDGCKAPTAPETSPN--EKPT---YQESTETHFKECVGAQY-----682

QY  221 AVPKGTDALVRVDFDFHGNAYLENYKVEIKPIPKLNOGTTTA-----GNKIP-- 271
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  683 -----QERIGDGAPEISCDVSKLHAPETPGASLESCISGCGHCNGKNSPSD 729

QY  272 -----VTFMANAYLDNQSTIVIVPILKXENOTDKPSILPQFKRNKAQEN-----SKL 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  730 ASTRETTSLGESKNENKIEALEVPCAD-EMQSOILQYEHFRNENIDKKAQVKVSKL 788

QY  320 DEKVEEPEKT-----SEKVEKEL-----SETGNSTSNSTLEEYPTDVPQVEKV 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  789 EESVEYETPNFQSSSTAHEGETETVEKEMFSFDELFPQNKNGITTEAPPESLIHKEI 848

QY  363 AKFA--ESYGMKLENVFNMDGTIE-----LYLPSSGEVIKKNM 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  849 KKFQTEEKAYITLGGDVVQKSGSLERANITLSASANEENEEAANAFVEGINVETHV 908

QY  399 ADFTGEAPQNGENKPKSENGKVSTGTVENOPTENKPADSLPEAPNEKPKVPKPSNDGML 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  909 STY-GTSVEDSDQIQDSENRMNMDGMDLVSHGNEEAARDPWLDNSEKVSQVEEIFSHEEGQL 967

QY  459 NPEGNVGSDDP 468
Db  968 SVEGGIDGGP 977

RESULT 5
Q9Z4N7
ID  Q9Z4N7 PRELIMINARY; PRT; 1873 AA.
AC  Q9Z4N7;
DT  01-MAY-1999 (TrEMBLrel. 10, Created)
DT  01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Surface protein precursor.
OS  Enterococcus faecalis (Streptococcus faecalis).
OC  Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
   OC  Enterococcaceae; Enterococcus.
OX  NCBI_TaxID=1351;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=MMH594;
RX  MEDLINE=99081742; PubMed=9864215;

```

RA Shankar V., Baghdayan A.S., Huycke M.M., Lindahl G., Gilmore M.S.;
 RT "Infection-derived Enterococcus faecalis strains are enriched in esp,
 RL a gene encoding a novel surface protein.";
 DR Infect Immun 67:193-200(1999).
 DR EMBL: AF034779; RAD09858.1; .
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR TIGRfam: TIGR01167; LPXTG_anchor; 1.
 DR TIGRfam: TIGR01168; YSRK_signal; 1.
 KW Signal.
 FT SIGNAL. 1 49 POTENTIAL.
 FT CHAIN 50 1873 SURFACE PROTEIN.
 FT SEQUENCE 1873 AA; 202084 MW; F609483DB9B80621 CRC64;
 Query Match 5.6%; Score: 153; DB 2; Length 1873;
 Best Local Similarity 23.7%; Pred. No. 10;
 Matches 123; Conservative 58; Mismatches 194; Indels 144; Gaps 28;
 QY 2 KLDKKI---EKKIAGIMKQYGVKRESIVVKNKNAIIPYHGDHHDADPIDEHKPVGIG 57
 Db 1027 KDLPKDAKYTWKEKVD--ISTAGNKKGTVVVT-----YSDGSDEVE-VD---VTVT 1072
 QY 58 HSHSNLELEKPE--EGVAKKEGKNKVTGELTNVNVLLKNSTFNN-----CNFTLANGQ 109
 Db 1073 DNRSDADKYEPTVEGKEVIGKGVLDLTONVNLPTLPQGTITVDTPGTDITNTPGNVE 1132
 QY 110 KRVSFSPPELEKKGILNMLVKL---ITPDGKVLKVKSGKVFEGV-----GNIAN 157
 Db 1133 GVIEVTPD-----GTRKDTVKVPVEVTNRSADKYEPTVEGKVEIGKGVLDLTDNVTN 1186
 QY 158 FELDQVLPQGTQKFTYASKDPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVN 217
 Db 1187 L-----PTLPQGT---TJT-----DVTGPGTIDTNPNGVEGIV-----TY---- 1221
 QY 218 POPAVPKGTDALVRVDFEFGNAYLENNY-----KV---GEIKLP-----IPKLNQGT 262
 Db 1222 -----PDGTRKDTVKVPVEVTNRSADKYEPTVEGKEVIGKGVLDLTONVNLPTLPQGT 1276
 QY 263 T---RTAGNKIPVTFMANAYLDNQSY-----IVEPFILEKENOTDKPSILPQFKRKA 313
 Db 1277 TVTVDTPGTTDNTFCNGEIVGIEVTPDGTQDTVKVPVEVTNRSADKYEPTVWVEGKV 1336
 QY 314 QENSKLDEKVEEPTSEKVEKEKLSGTSTNSTLEEYPTVDPVQEKVAKFAESYGMKL 373
 Db 1337 EIGKQVD-----LTDNVNLPTLPQGTITVDTPG-----GTID 1370
 QY 374 ENVLNMDGTIELYLPSPG--EVTK-----KNMADTGEAPQNGENKPSNGKYS-TGT 424
 Db 1371 TNPNGYEGVIEVTPDGTQDTVKVPVEVTNRSADKYEPTVEGE-KVEIGKGVLDLTDN 1429
 QY 425 VENQPTENKPADSLPAPNEKPKYPENSTNDNGMLNPEGN 463
 Db 1430 VTNLPT-----LPQGTITVDTPGTTIDT---NTPGN 1458
 RESULT 6
 Q9PQ08 PRELIMINARY; PRT: 4688 AA.
 ID Q9PQ08
 AC Q9PQ08;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein U0482.
 GN U0482.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Ureaplasma.
 OC NCBI_TaxID=134821;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROVAR 3;
 RX MEDLINE=20500219; PubMed=11048724;
 RA Glass J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,

RA Cassell G.H.;
 RT "The complete sequence of the mucosal pathogen Ureaplasma
 RL urealyticum.";
 DR Nature 407:757-762(2000).
 DR EMBL: AE002145; AAF30894.1; .
 DR InterPro: IPR001152; Thymosin_b4.
 DR SMART: SM00152; THI; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 4688 AA; 534880 MW; B53ABFAFEE1997E CRC64;
 Query Match 5.4%; Score 149.5; DB 16; Length 4688;
 Best Local Similarity 20.6%; Pred. No. 54;
 Matches 125; Conservative 89; Mismatches 219; Indels 173; Gaps 27;
 QY 1 MKDLDKIEKIAIGIMKQYGVKRESIVVKNKNAIIPYHGDHHDADPIDEHKPVGIGHSH 60
 Db 3859 INDQQTINVTLSGVNSKYN-GROIKVYKDNVNIY-----ESSLITLQKKG 3905
 QY 61 SNYELFPEEGVAKKEGKNKVTGELTNVNVLLKNSTFNNQNTLANGQKRVSFSPPEL 120
 Db 3906 NDYQLL-----LSNLNSNREYRFEKIE--INHISN-TNNEDELEKLG---VSNFTITQT 3954
 QY 121 EKKL-----GINMLVKLITPDGKVLKVKSGKVFEGVGNIANFELDQYLP 166
 Db 3955 KNTTVQWNSSATIVGTRGVNFNFKIKSED-KILEN-----NQGVVAMFAP 3999
 QY 167 QTEK-----YTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVN 219
 Db 4000 KETIRDTNWTQYTRPLKDVTSDFEGTWAHDLNSNVNFKET-----TYKLKVIQ 4050
 QY 220 FAVPKGTDALVRVDFEFGNAYLENNY-----KVGEIKL-PIPKLNQGTITET--- 265
 Db 4051 F-VNKPTKAKNNI--NSENNAVILDNTNSINSNYEFTTKVGDHKLINITSSNNVNTNSQTI 4108
 QY 266 -----AGNKIPVTFMANAYLDNQSYIVEPFILEKENOTDKPSILPQFKRKAQ 314
 Db 4109 NFTLSGVKKSQWVGKIKLSYKSN-----DTSESHTEVLESNKTKQTNILLNLKRNRTY 4164
 QY 315 ENSKLEKVEEPTSEKVEKEKLSGTSTNSTLEEYPTVDPVQEKVAKFAESYGMKL- 373
 Db 4165 -TLIDYKLDNNVNSDPPKEGNTNSFIITRTSAINVLEIEISNASTNLKSTIKIN 4222
 QY 374 -----ENVLFNMDGTIELYLPSPGSEVVKKNMADTGEAPQNGENKPSNGKYSTGT----- 424
 Db 4223 LNDPDNVLRLDQDQATIVY-----GNNKQAMGFTVSGNIKYLTATLVLDLN 4267
 QY 425 -----VENQPTENKPA-----DSLPEAPNEKPKYPENSTND---NGM 457
 Db 4268 FNDKVTNVIISFNKKPSTAAENIGDKSNIIYNDSIP-----KLEINNDIIVNGP 4319
 QY 458 LNPE-----GNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLGLD-----SVIFNMD 506
 Db 4320 INKEIVKVNANQKNNDVDLGLQINPKIAHLNRFIAKFKSTNNDIETNINGSLSLVND 4379
 QY 507 GTIELR 512
 Db 4380 GKTSIR 4385
 RESULT 7
 Q25662 PRELIMINARY; PRT: 1939 AA.
 ID Q25662
 AC Q25662;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Repeat organellar protein.
 OS Plasmodium chabaudi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCBI_TaxID=5825;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=96V;

RA Werner E.B., Holder A.A., Aszodi A., Taylor W.R.;
 RT "A novel 11-residue coiled-coil motif predicts a histidine zipper";
 RL protein Pept. Lett. 3:139-146(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-96V;
 RX MEDLINE-98418765; PubMed-9747969;
 RA Werner E.B., Taylor W.R., Holder A.A.;
 RT "A plasmidium chabaudi protein contains a repetitive region with a
 predicted spectrin-like structure";
 RL Mol. Biochem. Parasitol. 94:185-196(1998).
 DR EMBL: U43445; AAC63403.1;
 PR PRO1592; LCRVANTIGN.
 SQ SEQUENCE 1939 AA; 229002 MW; B36E462001C6F22F CRC64;

Query Match 5.4%; Score 148; DB 5; Length 1939;
 Best Local Similarity 19.3%; Pred. No. 20;
 Matches 117; Conservative 101; Mismatches 206; Indels 182; Gaps 27;

QY 4 LDKIEETAGIMKOYGVKRESIVVNKKNALIIYPHGHHH---ADPIDEHKPV----- 54
 Db 989 LGEEHEVAGIEEY--KVEAIKAEERKDVVTGKGQHEIAKLEDGCHKEVYVNEVEK 1046
 QY 55 -----GIGHSHS-----YELFKEEGVAKKEGKVYTGELTNV 90
 Db 1047 KVASLLNMLENHNKEMIKLEEHEKESADIVKLYQDEEV-KNSNNKI---BELTNVI 1102
 QY 91 NLLKNSTNNQFTILANGQKRVFSFPPELEK-KLGINMLVKLITPDGKVLKESG-KVF 148
 Db 1103 KDLNDSIMCYKQOILEEVEKREYN--BEINKIKIVQEMKDM--NDKKILEKENEIKKL 1158
 QY 149 GEGVGNIAFELDQPLPGQTFKYTIASKDYPEVSDGTFTVPSLAYKMASQTIFFPH 208
 Db 1159 NKKLSNYKVFETK-----NTYK----- 1176
 QY 209 AGDTLYRVNPOPAVPGKTDALV--RVDFEHGNAYLENNYKVGEIK--LPiPKLNOGTR 264
 Db 1177 -----NSEMVNENKRIIVDSVCHENISESDEG--KGNLKWTLSEKKERNIFS 1246
 QY 265 TAGNKIPVTFMANAYLDNQSYIVVPLEKENQTKPSI-----LPQF 308
 Db 1227 INDKNSESELVDTT--KSAYINKIEMKKEIEDNGKIEDLNKILDLSNELINLEM 1283
 QY 309 KRKAQENSKL-----DEKVEEPTSEKVE-----REKLETSNSTSLTE 350
 Db 1284 KNLVDENNLLKKEIEIKDKLNKEKKNENTEILNDDIILKKEISEWDEEKLTK 1343
 QY 351 EYPTVDVPOEVKAKFAESYGMKLENV--FN-----MDGTIELXLPSEGVKKNMAD 400
 Db 1344 NI-----KLKNDIEQINKEYKIKEENLMIKFNENIVTSLNQIEIEKMKLEELKNVEL 1399
 QY 401 FTGEAPOGNGENKPSKGVSTGTVENOPTENKPADSLPEAPNKPYPKPNSTNGMLNP 460
 Db 1400 LLAERETNMSISDNDKIVENNILED--TDSK-----QNNLNK---NV 1438
 QY 461 EGNVSDPMLDPALEAPVDPVQEKLEKFTASYGLGL-----DSVIFNMDG---TIEL 511
 Db 1439 EDKTGDDINCEKNNDQAKEISYLVKDEIKKISMLYGEELNRKNSYDEKVKNTLNEKLKI 1498
 QY 512 RLPSGE 517
 Db 1499 RNKKEG 1504

RESULT 8
 Q9NG79
 ID Q9NG79 PRELIMINARY; PRT: 2151 AA.
 AC Q9NG79;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE Hypothetical 232.4 kDa protein.
 OS Trichomonas vaginalis.

OC Eukaryota: Parabasalidea; Trichomonadida; Trichomonadidae;
 OC Trichomonas
 OX NCBI_TaxID=5722;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-#202;
 RA Delgaty K.L.;
 RT "The Molecular Identification and Characterisation of a Candidate Gene
 for Cell Detaching Factor from Trichomonas vaginalis";
 RL Thesis (2000), University of Ottawa, Ottawa, ON, Canada, In press.
 DR EMBL: AF257323; AAF67197.1;
 KW Hypothetical protein.
 SQ SEQUENCE 2151 AA; 232366 MW; F2EF91C5D9C60C82 CRC64;

Query Match 5.4%; Score 147; DB 5; Length 2151;
 Best Local Similarity 20.7%; Pred. No. 26;
 Matches 133; Conservative 90; Mismatches 197; Indels 222; Gaps 32;

QY 5 DKIEEKIAGIMKOYGVKRESIVVNKKNALIIYPHGHHHADPIDEHKP---VGIGHSHS 61
 Db 712 EEKPKLQGGI--KLGN SQPNQPLEKPKSGINLNLG--KSQPSSEKPKLGLNLKSPS 766
 QY 62 NYELFKPE---EGVAKKEGKVYTGELTNVNLKKNSTFNQNF-----TIANGQKRVSF 114
 Db 767 NSTEEKPKLGTGGSISLGNKPKQTEERPKLSLNLPKSPNQNSTEEKPKLQGLGLKLN 826
 QY 115 SFPPELEKK-----LGINMLVKLITPDGKVLKESGKVGEGVGNIAFELDQ 163
 Db 827 GNRQPTQTEERPKLQGLGKILNLSKSKQTEP-KPKFQLGGIKLGQSPN----- 876
 QY 164 YLPQTKYTIASKDYPEVSDGTFTVPSLAYKMASQTIFFPHAGDTLYRVNPOFAPV 223
 Db 877 -----STEEKPKLSLG-----TKLAQSP-----SNEKPKLSPN--LP 908
 QY 224 KGTDALRVDFEHGNAYLENNYK--VGEIKLPI-----PKLNOG-----TTR 264
 Db 909 KSP-----SNQTEKPKLQGLGKILNLSKSKQTEERPKLSEGGIKLGNVSSQ 957
 QY 265 TAGNK-----IPVTFMANAYLDNQ---STIVVEPIL-----EKENQTKPSI- 304
 Db 958 TSDEKPKLGGISLNFNGNQTEERPKLSSQNVKPTLGLGITLGOQTSEENKPKLG 1017
 QY 305 --LPQFKENKAQENSKL-----DEK-----VEEPTSEKVEKEL 337
 Db 1018 LNPV--KSNQTEERPKLGTGGSISLGNKPKQSEKPKLQGLGKILAHSPNQSTEEKPKL 1075
 QY 338 S-ETGNSNSTLEEVPTDVPQEKVAKFAESYGMKL-----ENVLFNMDGTIEL 387
 Db 1076 SLNLGKSPSNQSTEEKPKLG-----LGGLKLNGLHLSQDEKPKFSLGOMKLGK 1124
 QY 388 LPSGEVKKKNWADFTGEAPOGNGENKPSKGVSTG-----TVENQPTENKPADSLPEA 441
 Db 1125 VPSNE--SQEPPPKFTLNPPLPSNQAEQNDGKPKLGLGTSLSLNSQKEENEEKPKLGIA 1182
 QY 442 PNEKPVKPNSTNGMLNPEGNVSDPMLDPALEAPV--DPVQEKLEKFTASYGLGLDS 500
 Db 1183 P-----KPIQSNQN-----KIEKPKVINSVTKTEERKAISAPGVKLD- 1219
 QY 501 VIFNMDGTIELRP-----SGEVKKNLSDF 526
 Db 1220 -----LKLPKKGFGMLROPKPSLEVSNDSEQQNLSSF 1251

RESULT 9
 Q9USW1
 ID Q9USW1 PRELIMINARY; PRT: 791 AA.
 AC Q9USW1;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Hypothetical 85.5 kDa protein.
 GN SPCC21B10.03c.
 OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA McDougall R.C., Rajandream M.A., Barrell B.G., Skelton J.,
 RA Churcher C.M.,
 RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL121794; CAB57927.1;
 KW Hypothetical protein.
 SQ SEQUENCE 791 AA; 85523 MW; 02579D93CF5D68B8 CRC64;

Query Match 5.3%; Score 146.5; DB 3; Length 791;
 Best Local Similarity 22.1%; Pred. No. 7.1;
 Matches 115; Conservative 67; Mismatches 183; Indels 155; Gaps 26;
 QY 46 DPIDEKPY-GIGHSHSYELFKPEGVAKEGKVTGEEELTNVNNLLKSTFNQNT 104
 DB 126 DPEDAGVPLAGL-----EESTDNVEWDOFATNEKLFV-----KSHFEDLYT 168
 QY 105 LANGQKRVSEFPPELEKGLGINMLVK---LITPDGKVLKVKSGVFGEV-----G 153
 DB 169 -----SRIDRSHPKYKEKEQEDRAKEIEGVTNIIHAEERGLKVDSDGDEEDLYSG 223
 QY 154 NIANFELDOPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTY 213
 DB 224 VHRSIDVVRNYT--RSNAYNNKQKQPKNHEAPHQP-----QKVVVP---DDPA 270
 QY 214 LRVNPOFANPK-GTD--ALRVVFDEEHGNAYLENNYKVEIK-----LPIKLNQ 261
 DB 271 IVSHRHALLPRAPGPDRAAEAFRRNARAGPLSRREKDGQKEFMQFSQSLKIGSLD-- 328
 QY 262 TTRTAGNKLPVTFMANAYLDNOSTYIVPEILEKGNQTDKPSILPQFKRNKAQENSKLDE 321
 DB 329 -----SKQP-----SSTKSAEVKVADEKQLDPASS-----QATPADSREPRKE 367
 QY 322 KVEEPT-----SEKVEKEKLGSTNSTSTLEEVTPDVQEK-VAKFAESYGMKL- 373
 DB 368 EAEKPYTATEVSERKEV-----VDGNTSPSKEEKSTPEKPSVVTQKETTGTGLG 423
 QY 374 -----ENVLFNDGTTLELPSGEVKKKNADFTGEAPQ----- 407
 DB 424 TKLNKAIEFKPNVAPVTPGKFTIPSPAPVNASRPMQPSNNSEASIPSTTPQSPS 483
 QY 408 --GNGENKPSSE-----NGKYST-----GTVENQPTENKPAD-----SLPEAPNEK 445
 DB 484 VVNGENKPSSESPVFPFNGPVSSKEKEPILDNFNVKNGVBEHQAEQIDKPFSCPTWTNG 543
 QY 446 PVKPNSTDNGLN--PEGNVGSDPMLDPALEAPAVDPV 483
 DB 544 P-----LSLQQTANSRPEGNSGS-----AKKAAANPM 572

RESULT 10
 Q98PR8 PRELIMINARY; PRT; 1272 AA.
 ID Q98PR8
 AC Q98PR8;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE Hypothetical protein MYPV_6510.
 GN MYPV_6510
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UAB CTIP;
 RX MEDLINE=21267155; PubMed=11353084;
 RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Callisson F.,

RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis";
 RL Nucleic Acids Res. 29:2145-2153(2001).
 DR EMBL; AL445565; CAC13824.1;
 DR MYPULIST; MYPV_6510;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1272 AA; 147169 MW; 90019B854E6C6D74 CRC64;
 Query Match 5.3%; Score 146; DB 16; Length 1272;
 Best Local Similarity 20.3%; Pred. No. 14;
 Matches 125; Conservative 94; Mismatches 238; Indels 158; Gaps 28;
 QY 1 MKDUDKIEEKIAGIKQYGVKRESIVVNRKNAIIVPHGDHHDHDPIDHKPVGIGHSH 60
 DB 86 LTDFTDQLDEKINVSYK-----LISNEQK-----QEENSQRNSQN 121
 QY 61 SNYELFPPEGVAKEGKVTGEEELTNV-----NLLKNSTFNQNTLANGQKRVSP 114
 DB 122 QNEKFANDDEGLTKVLVEFSYNGVPTREYKLGKFKNTIQKTSNNLNLNLSNKEQ----- 177
 QY 115 SFPEL-----EKLGINMLVK---LITPDGKVLKVKSGVFGEVGNIANFELDQPLPGQ 168
 DB 178 IHPSELDSQQQLVTTNFELKSLLELTDIEKYVLVEKVVVNDIDAEAKIRLTQLKTNE 237
 QY 169 TFK--YTIASKDYPEVSYDGT-----TVPTSLAYKMASQTIYFPFHAGDTYLRNP-- 218
 DB 238 NIKNSFEVLKDKFKYKNDLSSKAKINLVKNDLSIELNITSKGFANKTDEDDVKNLK 297
 QY 219 --QFAPKGTDALVRVDFEHGNAYLENNYKVEI--KLPIKLNQGTTRTAGNKIPVT 273
 DB 298 ETLNIDLPNGVEF-----EFVSLAPKANDASVGLLYTKLVKNNLNGTSENPSK-- 347
 QY 274 FMANAYLDNOSTYIVPEY--ILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEK 331
 DB 348 --NGRIET-NTIVEETNLNLSKNDKDEP-----KDNSTNSKDEPKTDEKVEEP 397
 QY 332 VEKE-KLSEGTGNTSNTLEE-----VPTVDVPQEKVA----- 363
 DB 398 REDEPKTNMPSDKDKPKIDEPNEPKDQKTEPKNEPKPKPKVEPADETLAIFDKISKI 457
 QY 364 -----KFAESYGMKLEN--VLFNMDGT--IELYLPSEGEVKKKNADFTGEA 405
 DB 458 ELKENSQKLPQSPKES-DLNLNLKVLVSDDKNKFSELSLPQGYSTFSLA----- 510
 QY 406 PQNGENKPSNGKVSSTGV-----ENQPTENKPADSLPEAPNEKPKPENSTDNGLN 459
 DB 511 -----SNSNDEGTLVKVIVQKQKVKTKELKLTLLTEFESLK-----ESDFQLDFSN 561
 QY 460 PEGNVGSDPMLDPALEAPAVDPVQEK-----LEKFTASYGL-GLDSVIFNMDGTIELR 512
 DB 562 KKDRIASSVVNDKIIKESLV--VKNTIENFDNFKNIDISYSSSLDEV---NGKLKIK 615
 QY 513 LPSGEVKKNLSDFT 527
 DB 616 MTIFKTKDRLKEEL 630
 RESULT 11
 Q963T1 PRELIMINARY; PRT; 1078 AA.
 ID Q963T1
 AC Q963T1;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Glutamate-rich protein (fragment).
 GN GLORP.
 OS plasmodium reichenowi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5854;
 RN [1]
 RP SEQUENCE FROM N.A.


```
RX MEDLINE=21313688; PubMed=11420113;
RA Thelsen M., Thomas A.W., Jepsen S.;
RT "Cloning, nucleotide sequencing and analysis of the gene encoding the
RL glutamate-rich protein (GLURP) from Plasmodium reichenowi.";
RL Mol. Biochem. Parasitol. 115:269-273(2001).
DR EMBL; AF356828; AAK40236.1;
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 1078 AA; 123581 MW; 80388088B223913C CRC64;

Query Match      5.3%; Score 145.5; DB 5; Length 1078;
Best Local Similarity 20.0%; Pred. No. 12;
Matches 111; Conservative 74; Mismatches 176; Indels 195; Gaps 23;

QY 2 KDL-----DKLEEKIAGIMQYGVKRESIVVNKKNALIIYPHGDHHDADIDHKPVGI 56
DB 161 KDLQDLDLEPPPTQIHKDY--KEENLI--NEEDSEPPFRHEKKLDNHYEKNVHF 216
QY 57 GHS--HSNYELFKPEE-----GVAKKE-----GNKYVTGEELTNVNVNLLKNSTFN 99
DB 217 ENSVNSNQENLKPESFDEHLNINSKLEDOLDLHEHCENSEHLKDE-----KIGNEPLV 270
QY 100 NONFTLANGQKRVSPFPPELEKLGIGINLVKLITPDGKVLKESGKVGEGVGNIANFE 159
DB 271 HENLSPNDPIEQIILNLEPQ-----ETNVQQLYNEKQNVKEKPNSEIFSLDLKETTNE 325
QY 160 LDOPVLPQGTETKTIASDYPEVSYDGTFTVPTSLAYKKASQTIYPFHAGDTVLRVNPQ 219
DB 326 I----LPNQ-----NPL 333
QY 220 FAVPKGTDALVRVDFEFGNAYLENNYKVGKIPKLNQGTTRTAGNKIPVTFMANAY 279
DB 334 ENIKOSESEINHVDH-----ALPEN-----IIDK 359
QY 280 LDNSTYIVE---VPILEKENQDKPSILPQFKEN--KAQENSKLDEKVEBPKTSERKEV 334
DB 360 LDNEKEHIDESHQNVNVLQ--ENNINNHOLEPQERANIESFEPKIDSEIILP-----ENVEK 415
QY 335 EKL-----SETGNSTNSTLEPYTPDVOEKVAKFAESYGMKLENVLF 378
DB 416 BEIYVDVPSPKHFHNETLEQETSESEHEAVSEKNAHETVHEBAVSQESNPERADN-- 472
QY 379 NWDGTIELYLSGSEVIKKNMADFTGEAPQNGENKPSKNGKSTGTVENQPTENKPADSL 438
DB 473 --DGNV-----SONSNELNENEFVESEKSEHEFAENES--SL 507
QY 439 PEAPNEKPKPENSDNG-----MLNPEGNVGSDPMLDPALEAPAVDFVQEKLEKETASY 494
DB 508 EEAHQEEIVPQNQNGESGSKLVNDGE-----GFEAHQEEIVPQNQNGES--- 554
QY 495 GLGLDSVIFNMDGTIE 510
DB 555 --GSKLVNDNDEGGFE 568

RESULT 12
Q54875 PRELIMINARY; PRT; 1927 AA.
ID Q54875;
AC Q54875;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Iga1 protease.
DE IGA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PK81.
RX MEDLINE=97047666; PubMed=8926055;
RA Poulsen K., Reinholdt J., Kilian M.;
```

```
RT "Characterization of the Streptococcus pneumoniae immunoglobulin A1
RT protease gene (iga) and its translation product.";
RL Infect. Immun. 64:3957-3966(1996).
DR EMBL; X94909; CAA64396.1;
DR MEROPS; M26.001;
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMS; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMS; TIGR01168; YSIK_signal; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 1927 AA; 215170 MW; BAF74DB811301037 CRC64;

Query Match      5.3%; Score 145.5; DB 2; Length 1927;
Best Local Similarity 22.4%; Pred. No. 27;
Matches 126; Conservative 79; Mismatches 207; Indels 151; Gaps 34;

QY 3 DLDKKIEKTAGIMQYGVKRESIVVNKKNALIIYPHGDH--HHADPIDEHKP--VGIGH 58
DB 197 DFEVSNQEKSAAT-----PTKQKQVDVNVTFNFDHPSSTVQAIQEQTPVSSTKPTQVVE 252
QY 59 SHSNYELFKPEEGVAKKEGNKVTGELTNVNVNLLKNSTENNQNFNLANGKRVSEFP 118
DB 253 KPFSTELINPR-----KEEKSSDSQEQLAPEHKNLEYK-----EEKIS--PK 293
QY 119 ELEKKGINMLVLIITPDGKVLKESGKVGEGVGNIANFELDPVLPQGTETKTIASKD 178
DB 294 E---KTGVNLT---NPQDEVLS-----GOLNK----- 314
QY 179 YPEVSYDGTFTVPTSLAYKKASQTIYPFHAGDTVLRVNPQFAVPKGTDALVRVDFEFG 238
DB 315 -PELLYREE-TIETKIDFQEEIQE--NPDLAECT--VRKQEGKLGKRV--IVRIFS--- 364
QY 239 NAYLENNYKVG-----EIKLPIKL-NQGTTRTAGNK-IPVTFMANAYLDNSTYIVEY 290
DB 365 ---VNKEEVSREIVSTSTTAPSPRIVEKTKTKVIKEQPETGV--EHKDVQSGAIVE- 417
QY 291 PILEKE-----NOTDKPSTLPQFKRKAQENSK--LDEKVEBPKTSERKEKLEKLTGN 342
DB 418 PAIQPELPEAVVSKGEPEVQPTLPFAVVTDKGEPVQPELPEAVVSDKGEPEQVAPLPE 477
QY 343 STNSLTIEEYTPDVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEV-IKKNMADF 401
DB 478 YKGN--IEQVKPPEPVE---KTEQGEKPEEVPVK-----PTEETPVNPEGIT 522
QY 402 TGEAPQNGEN--KPSKNGKSTGTVENQPT--ENKPADSLPEA-----PNERKVPK 449
DB 523 EGTSTIQG-AENPVQPAEDTQTNKIANENTGEVSNKPSDSKPPVEESNQPKNGTATRP 581
QY 450 ENS-----TDNGMLNPE---GNVSGSDPMLDPALEAPAVDFVQEKLEKFTASYGLGDSVI 502
DB 582 ENSGNTTSENGQTEPEPSNGNSTEDVSTKSTNSNGNEIKQENE-----LDPDKKV 634
QY 503 FNMDGTIELPSPGEVIKKNLSD 525
DB 635 EDPEKLTLELR-----NVSD 648

RESULT 13
O96134 PRELIMINARY; PRT; 2485 AA.
ID O96134;
AC O96134;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Ser/Thr protein kinase.
DE PF00150C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
```

RESULT 14	
Q9CFL1	
ID	QCFL1
AC	QCFL1;
DT	01-JUN-2001
DT	01-JUN-2001
DT	01-JUN-2002
DT	01-JUN-2002

PRELIMINARY: PRT: 1441 AA.

{TREMBlrel. 17, Created}
{TREMBlrel. 17, Last sequence update}
{TREMBlrel. 21, Last annotation update}

RESULT 15

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
RN NCB1_TaxID=1311;
RP [1]
RC SEQUENCE FROM N.A.
RX STRAIN=SB35;
RA MEDLINE=91257158; PubMed=2044657;
RA Haden L., Frithz E., Lindahl G.;
RT "Molecular characterization of an Iga receptor from group B
RT streptococci: sequence of the gene, identification of a proline-rich
RT region with unique structure and isolation of N-terminal fragments
RT with Iga-binding capacity.";
RL Eur. J. Immunol. 20:2241-2247(1990).
RN Eur. J. Immunol. 21:1481-1490(1991).
RP [2]
RC SEQUENCE OF 39-56.
RX STRAIN=SB35;
RA MEDLINE=91055597; PubMed=2242758;
RA Lindahl G., Akerstrom B., Vaerman J.-P., Stenberg L.;
RT "Characterization of an Iga receptor from group B streptococci:
RT specificity for serum Iga.";
RL Eur. J. Immunol. 20:2241-2247(1990).
CC !- FUNCTION: PROTEIN BAC, IS A SURFACE PROTEIN WITH BOTH IGA-BINDING
CC !- CAPABILITY AND A REGION CORRESPONDING TO THE BETA ANTIGEN.
DR EMBL; X58470; CAA41384.1; .
DR PIR; A60234; A60234.
DR InterPro; IPR004829; Csurface_antigen.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003599; Ig.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR ProDom; PD153432; Csurface_antigen; 1.
DR SMART; SM00409; Ig; 1.
DR TIGRFAMS; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMS; TIGR01168; YSIK signal; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
KW Receptor; Signal; Transmembrane; Repeat; Cell wall; Virulence.
FT SIGNAL 1 38
FT CHAIN 39 1134 IMMUNOGLOBULIN ALPHA FC RECEPTOR.
FT DOMAIN 1002 1007 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
FT TRANSMEM 1111 1129 MEMBRANE ANCHOR REGION (BY SIMILARITY).
FT REPEAT 826 915 30 TRIPEPTIDE REPEATS (MOTIF XP2).
SQ SEQUENCE 1134 AA; 127980 MW; FC10A200B3DB7CE1 CRC64;

Query Watch 5.1%; Score 141; DB 2; Length 1134;
Best Local Similarity 20.0%; Pred. No. 22;
Matches 126; Conservative 95; Mismatches 207; Indels 202; Gaps 32;

QY 2 KLDKKIEKI-----AGIMK--QYGVRESIVNKEKNAITYPHGDHHADPI--- 48
Db 240 EDAEVKVRRELKLFSSSTAGLCQEIQEHVKRET---SSEENT---QKVDEHYANSLQL 293
QY 49 -----DEHKPYGIGHSHSNYELFKPEGSVAKKEGNKVY--TGEELTNVNVLL 93
Db 294 AOKSLELDKATTNEGATQVKNOFLNACKLKEIQLIKETNVKLYKAMSELEQVEKEL 353
QY 94 KNSTFNNOFTLANGOKRYSFSPFLEKLGINMLVLIITPDGKVLKVSOKVFGVG 153
Db 354 KHNSEANLEDLVAKSKEIVR-----EYEGKL--NOSKNL--PELKQLEEEAHSKLKQVVE 404
QY 154 NI-ANFELDQPYLPQGTFFKYTIASKD-----YPE--VSYDG--TFTVPTSLAVK 197
Db 405 DFRKKFKTSEQVTPKKRVKRDLAENNOQKIELVSPENITVIEGDEVKFTVTA---K 460
QY 198 MASQIFYPFHAGDYLRYVNPQFVAVPKGTDALRVRFDEFHGNAYLENNYKVEIKLPKP 257
Db 461 SDSKT---TLDFSDDLTKYNPSVS-----DRISTNYKNT---DNHKAETIKNLK 506
QY 258 LNOGTRT-----AGNKIPVTFMANAVLDNGSTVIVEPILKEKNTQDKPSILPQFKN 311
Db 507 LNESQTVTLKAKDDSGNVVEKTF-----TITVQKKEK-----QVPKT 544
QY 312 KAQENSKLDEKV-EEPFKTEKVEKEKLSETGN-----STNSSTLEEVPV----- 355

```

```

Db 545 PQKDSKTEEKVPQEPFKSNDKNQLOELIKSAQQELEKLEKAIKELMEQPEIPSPNPEYGIQ 604
QY 356 -----DPQOEKVAKP-----ABSXGMKLENVLFNMDGTIELYLPSPGEVIK 395
Db 605 KSINESQKEPIQEAITSFKKIIGDSSSKYYEYHYNKTKSDFMNQLHAQM-----EMLT 659
QY 396 KNMADFTGEAPQNGENKPSNGKYST-----GTVENQPT-----NK----- 433
Db 660 RKVVQYMKRYPDPAETIKKIFESDMKRTKEDNYGSLNDALKGYPEKYEFTLPFNKIKQIVD 719
QY 434 PADSLPEAPNEKPKVPKPSNSTONGMLNPGNNGVSDPMLDPALEEA-----PAVDPVQ 484
Db 720 DLDKKVEQDQAPI-PENSE-----MDQAKEKAKIAVSKYMSKVLDCGVH 762
QY 485 EKLEKFTASYGLGL-----DSVIFNMD 506
Db 763 QHLQKKNHSHKIVDLFKLELEAIKQQTIFDID 792

Search completed: May 13, 2003, 13:56:16
Job time : 38.447 secs

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:52 ; Search time 8.18433 seconds
(without alignments)
1898.176 Million cell updates/sec

Title: US-09-471-255-10

Perfect score: 2146

Sequence: 1 MKDLKKIEBKIAKIMKQY.....IELRPSGEVIRKNSLDFIA 528

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	184	6.7	571	US-08-961-083-4	Sequence 4, Appli
2	169.5	6.2	1964	US-08-790-912-3	Sequence 3, Appli
3	169.5	6.2	2052	US-08-790-912-2	Sequence 2, Appli
4	151	5.5	278	US-08-961-083-94	Sequence 94, Appli
5	149	5.4	1848	US-08-296-791-6	Sequence 6, Appli
6	149	5.4	1848	PCT-US95-10661A-6	Sequence 6, Appli
7	143	5.2	1164	US-08-923-992A-10	Sequence 10, Appli
8	141	5.1	1098	US-08-923-992A-8	Sequence 8, Appli
9	141	5.1	1164	US-08-923-992A-2	Sequence 2, Appli
10	140	5.1	1128	US-08-923-992A-6	Sequence 6, Appli
11	139	5.1	1104	US-08-923-992A-4	Sequence 4, Appli
12	138.5	5.0	984	US-08-242-932-2	Sequence 2, Appli
13	138.5	5.0	984	US-08-714-481-2	Sequence 2, Appli
14	138.5	5.0	984	PCT-US95-08111-2	Sequence 2, Appli
15	132.5	4.8	1164	US-08-589-756-1	Sequence 1, Appli
16	132.5	4.8	1164	US-09-206-800-1	Sequence 1, Appli
17	132.5	4.8	1164	US-09-206-898-1	Sequence 1, Appli
18	132	4.8	1231	US-08-904-031A-4	Sequence 4, Appli
19	130.5	4.8	1183	US-08-447-031A-2	Sequence 2, Appli
20	130	4.7	1702	US-08-296-791-5	Sequence 5, Appli
21	130	4.7	1702	PCT-US95-10661A-5	Sequence 5, Appli
22	129	4.7	10182	US-09-134-001C-3159	Sequence 3159, Ap
23	128.5	4.7	933	US-08-293-728-2	Sequence 2, Appli
24	128.5	4.7	933	US-09-421-868-2	Sequence 2, Appli
25	126.5	4.6	932	US-09-071-035-416	Sequence 416, App
26	126.5	4.6	969	US-09-071-035-414	Sequence 414, App
27	126.5	4.6	1167	US-08-589-756-2	Sequence 2, Appli

Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 7, Appli
Sequence 3530, Ap
Sequence 5080, Ap
Sequence 23, Appli
Sequence 66, Appli
Sequence 246, App
Sequence 250, App
Sequence 254, App
Sequence 470, App
Sequence 474, App
Sequence 478, App
Sequence 23, Appli
Sequence 4, Appli
Sequence 152, App
Sequence 22, Appli

ALIGNMENTS

RESULT 1
US-08-961-083-4
; Sequence 4, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 571 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-961-083-4

Query Match 6.7%; Score 184; DB 4; Length 571;

Best Local Similarity 23.5%; Pred. No. 6e-07;

Matches 130; Conservative 70; Mismatches 162; Indels 192; Gaps 34;

QY 60 HSNVELEKPEGVAKKEGKNVYT-----GELTNVNV-----LLKNSTFN-----NQN 102

Db 119 HKNLET-KREKISPKRKTGVNTLNPQDEVLSGLINKPELLYREETMETKIDFQEEIQEN 177

QY 103 FTLANGOKRVSEFPPELEKKGINM-LVKLIT-----PDGKVLKYS 144
DB 178 PDLAEGTVRV-----KQEGKLGKKEIVFISVKNKEVSREIVSTSTAPSRIVEKGT 231
QY 145 GK--VFGE-----GNIANFELDQYLPFGQTFKYTIASKDYPEVSDGTFTVP 191
DB 232 KKTQVIKEQPTGVHKGVSOGAIVEPAI-QPELP-----EAVSDKGPEVQP---TLP 282
QY 192 TSLAYKMASOTIYFPFHAGDYLVRNPQ-----FAVPKGTDALVRVDFHGNAYLENNYK 247
DB 283 EAVV-----TDKGET--EVQPESTVVSVDKGEPEQVAPLPEYKGN----- 321
QY 248 VGEIK--LPIPKL-NOGTRTAGNKIPVTFMANAYLDNQSTIYIVEVPILEKENTDKPSI 304
DB 322 IEQVKPPTVEKTEKQGEPT--EEVPV-----KPTETVPNPNNEGTEGTSI 367
QY 305 LPQFKN---KAOENSKLDEKVEPKTSEK-----VEKEKLSGTSNS 343
DB 368 --QEAENPVQPAEESTTNSKRV-SPDTSKNTGGEVSSNPSTSTSVGESNKPENHDSKNE 424
QY 344 TSNSTLEEVPYDVPQVKVAKFAESYGMKLENLNFMDGTIELVLPSEVIVKKNMADFTG 403
DB 425 NSEKVEEVP-VNP-----NEGTV-----GTSNOET 450
QY 404 EAPQNGENKPSNGKV---STGTVENOPTENKP--ADSLPEAPNEKVPKPKENS---TD 454
DB 451 EKPQPAEETQNSGKIANENTGEVSNKPSDSKPPVEESNQPKNGATKPKENSGNTTSE 510
QY 455 NGMLNPE---GNVSDPMDLPALEAPAVDPVQEKLEKFTASYGGLDVSIFNMDGTIEL 511
DB 511 NGQTEPEPSNGSTEDVSTESNWSNGNEIKQENE-----LDPKKVVEEPEKTIEL 563
QY 512 RLPSGEVIVKKNLSD 525
DB 564 R-----NVSD 568

RESULT 2

US-08-790-912-3
; Sequence 3, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020

TELEFAX: (215) 567-2991
TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1964 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-790-912-3
Query Match 6.2%; Score 169.5; DB 2; Length 1964;
Best Local Similarity 23.7%; Pred. No. 5.9e-05;
Matches 115; Conservative 60; Mismatches 140; Indels 171; Gaps 30;
QY 60 HSNYELFKPEEGVAKKEGNKYT---GEELTNVNV---LLKNSTFN-----NON 102
DB 279 HKNLET-KKEEKISPKREKTVNTLNPDQDEVLSGOLNKPPELLYREETIETKIDFQBEIQEN 337
QY 103 FTLANGOKRVSEFPPELEKKGINM-LVKLIT-----PDGKVLKYS 144
DB 338 PDLAEGTVRV-----KQEGKLGKKEIVFISVKNKEVSREIVSTSTAPSRIVEKGT 391
QY 145 GK--VFGE-----GNIANFELDQYLPFGQTFKYTIASKDYPEVSDGTFTVP 191
DB 392 KKTQVIKEQPTGVHKGVSOGAIVEPAI-QPELP-----EAVSDKGPEVQP---TLP 442
QY 192 TSLAYKMASOTIYFPFHAGDYLVRNPQ-----FAVPKGTDALVRVDFHGNAYLENNYK 247
DB 443 EAVV-----TDKGET--EVQPESTVVSVDKGEPEQVAPLPEYKGN----- 481
QY 248 VGEIK--LPIPKL-NOGTRTAGNKIPVTFMANAYLDNQSTIYIVEVPILEKENTDKPSI 304
DB 482 IEQVKPPTVEKTEKQGEPT--EEVPV-----KPTETVPNPNNEGTEGTSI 527
QY 305 LPQFKN---KAOENSKLDEKVEPKTSEK-----VEKEKLSGTSNS 344
DB 528 --QEAENPVQPAEESTTNSKRV-SPDTSKNTGGEVSSNPSTSTSVGESNKPENHDSKNE 585
QY 345 TSNSTLEEVPYDVPQVKVAKFAESYGMKLENLNFMDGTIELVLPSEVIVKKNMADFTGE 404
DB 586 SEKTVVEVP-VNP-----NEGTV-----CTSNOETE 611
QY 405 APOQNGENKPSNGKV---STGTVENOPTENKP--ADSLPEAPNEKVPKPKENS---TDN 455
DB 612 KPQVPAEETQNSGKIANENTGEVSNKPSDSKPPVEESNQPKNGATKPKENSGNTTSEN 671
QY 455 GMLNPE 461
DB 672 GQTEPE 677

RESULT 3

US-08-790-912-2
; Sequence 2, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-790-912-2

Query Match          6.2%  Score 169.5;  DB 2;  Length 2052;
Best Local Similarity 23.7%  Pred. No. 6.3e-05;
Matches 115;  Conservative 60;  Mismatches 140;  Indels 171;  Gaps 30;

QY  60 HSNVLELPKESGVAKKGNKYVT-----GEELTNVNV-----LLKNSTN-----NQN 102
DB  352 HKNLET-KKEKIKSPKKTGNTLNPDQVLSGOLNPELLYREETIETKIDFQEEIQEN 410
QY  103 FTLANGKRVSFSPPELEKKLGIM-LVKLIT-----PDGKVLKVS 144
DB  411 PDLARGTVRV-----KQEGKLGKVKVILRVFSVNKEVSRVISTTAPSPRIVEKGT 464
QY  145 GK--VFGE-----GV-----GNIANFELDQYLPGLQTFKYTIASKDYPEVSYDGTFTVP 191
DB  465 KKTQVIKEQPTGVHEKDVQSGAIVEPAI-QPELP-----EAVVSDKGEPEVQP---TLP 515
QY  192 TSLAYKMASQIFVFFHAGDYILRVNQ----FAVPKGTDALRVVDFEFGNAYLENLYK 247
DB  516 EAVV-----TDKGET--EVQPESDTVVSDKGGEQVAPLFEYKGN-----554
QY  248 VGEIK--LPFKL-NQGTTRTAGNKIPVTFMANAYLDNQSYIIVEVPILEKENOTDKPSI 304
DB  555 IEQVKPTEPVKTEKQEGPEKT--EEVPV-----KPTETPVNPNNEGTEGTSI 600
QY  305 LPQPKRN---KAQENSKLDEKVEEPKTEKV-----EKEKLSGTSNST 344
DB  601 --QEAENPVQPAEESTTNSEKVSPTDTSSENTGEVSSNPDSSTTSVGSNKPHEHNDKSNEN 658
QY  345 SNSTLEEPTVDVPQVEKFAESYGNKLENVLFNMDGTIELYLPSPGEVYKKNWADFTGE 404
DB  659 SEKTVEEVP-VNP-----NCTVE-----GTSNQETE 684
QY  405 APQNGENKPKSENGKV-----STGTVENQPTENKP--AOSLPEAPNEKPVKPKNS-----TDN 455
DB  685 KPVQPAEETQNSKIANENTGEVSNKPSDSKPPVBSNQPEKNGTKATKPKNSGNTTSEN 744
QY  456 GMLNPE 461
DB  745 GQTEPE 750

RESULT 4
US-08-961-083-94
; Sequence 94, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-94

Query Match          5.5%  Score 151;  DB 4;  Length 278;
Best Local Similarity 23.7%  Pred. No. 0.00011;
Matches 77;  Conservative 44;  Mismatches 102;  Indels 102;  Gaps 17;

QY  224 KGTCLALRVVDFEFGNAYLENNYKVGGEIK--LPFKL-NQGTTRTAGNKIPVTFMANAYL 280
DB  19 KGEPEQVAPLPEYKGN-----IEQVRPETVEKTKQEGPEKT--EEVPV-----60
QY  281 DNGSTVIVEPILEKENQTDKPSILPQKRN---KAQENSKLDEKVEEPKTEK-----331
DB  61 ---KPTETPVNPNNEGTEGTSI--QEAENPVQPAEESTTNSEKV-SPDTSSKNTGEVS 113
QY  332 -----VEKEKLSGTSNSTLEEVYPTVDVPQVEKFAESYGNKLENVLEN 379
DB  114 SNPSDSTTSVGSNKPHEHNDKSNSEKTVVEEVP-VNP-----150
QY  380 MDGTIELYLPSPGEVYKKNWADFTGEAPQNGENKPKSENGKV---STGTVENQPTENKP--434
DB  151 NCTVE-----GTSNQETEKPVPQPAEETQNSKIANENTGEVSNKPSDSKPPV 199
QY  435 ADSLPEAPNEKPVKPKNS-----TDNGLNPE---GNVSGDPMLDPALEAPAVDPVQVEKL 487
DB  200 EESNQPEKNGTKATKPKNSGNTTSENGQTEPEPSNGNSTEDVSTESNTSNGNEEIKQEN 259
QY  488 EKTASVGLGLSDVIFNMDGTIELR 512
DB  260 E-----LDPDKKVEPEKTLLELR 277

RESULT 5
US-08-296-791-6
; Sequence 6, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein

```

NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1389
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-6

Query Match 5.4%; Score 149; DB 4; Length 1848;
Best Local Similarity 20.1%; Pred. No. 0.0026;
Matches 117; Conservative 73; Mismatches 215; Indels 178; Gaps 26;
QY 28 VNKEKNAIYPHGHHADPTDEHKPVGIG-----HSHSNYELFKPEGVAKK----- 75
DB 704 LNVEKGLFLSGRTPPHARDI-----AGISSTKKDPHTENNEVVVEDDWINRNFRAITM 758
QY 76 --EGN-KVYTGEELTNVNNLLKSNFTNNQNTLANGOKRVSFSPPELEKKGILNMLVK- 131
DB 759 NVTGNASLYSGR---NVANITSNITASN-----NAQVHIGY-----KTGDTVCVRS 801
QY 132 -----LITPDGKLVKYSKGVFEGV-GNI-----ANFELDQYPLPGQTKYTIASKDYP 180
DB 802 DYTGYVTCHNSNLSEKALNSFNPTNLRGNVNLNENASFTLGKANLFG-----TIQSIGTS 856
QY 181 EVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 240
DB 857 QVN-----LKNSHWHLTGNSNVNQLNLTNGHILHNAQNDANKVTTNTLTNLSLNG 910
QY 241 -----YLENNYKVGEEKLPI-----PKLNOQT-----TRTAGNKIPVTM 275
DB 911 SFYVWVDFTNKSNKVVVKNKSATGNFTLQVADKGTGEPNHNLTLFDASNATRNNEVT-L 969
QY 276 ANAYLDN-----QSTYIVVEPILEKENOT-----DKPSILPOQKRN 311
DB 970 ANGSVDRGAWKYKLRNVNGRYDLNPEVEKRNQIVDTNITTPNDIQADAPS----- 1021
QY 312 KAGENSKLEKVEEP-----RTSEKVEKEKISE 339
DB 1022 -AQSNNEEARVETPPVPAPATASIASEQPEPTRAETAQAPAMEETNTANSTETAPKSD 1080
QY 340 TGNSTNSLTLEEVTVDVPOEVKAFSAESYGMKLENLVENMDGTTIELYLPSCSEVKKWA 399
DB 1081 TATQTNPNSESVPs-----ETTEKVAENPPQENETVAKNEQATEPTPQNGEVAKEQDP 1135
QY 400 DF-----TGEAPQNGENKPSKNGKYSTGVENQPT-----ENKPADSLPEAPNEKPV- 447
DB 1136 TVEANTQTNATQSGKTEQTQ-----TAEKSEPTSTVTSVSENPKEKTVSGSTEDKVV 1190
QY 448 ----KPNSTNDGMLNPEGNVSGDP-MLDPALFEAPVDPVQE 485

DB 1191 EKEERAKVETEETQAKQVTSKEPPKQAEPAPEEVPVTDINADE 1233
RESULT 6
PCT-US95-10661A-6
Sequence 6, Application PC/TUS9510661A
GENERAL INFORMATION:
APPLICANT: Washington University, et al.
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 18-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: FP-59941/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1389
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-6

Query Match 5.4%; Score 149; DB 5; Length 1848;
Best Local Similarity 20.1%; Pred. No. 0.0026;
Matches 117; Conservative 73; Mismatches 215; Indels 178; Gaps 26;
QY 28 VNKEKNAIYPHGHHADPTDEHKPVGIG-----HSHSNYELFKPEGVAKK----- 75
DB 704 LNVEKGLFLSGRTPPHARDI-----AGISSTKKDPHTENNEVVVEDDWINRNFRAITM 758
QY 76 --EGN-KVYTGEELTNVNNLLKSNFTNNQNTLANGOKRVSFSPPELEKKGILNMLVK- 131
DB 759 NVTGNASLYSGR---NVANITSNITASN-----NAQVHIGY-----KTGDTVCVRS 801
QY 132 -----LITPDGKLVKYSKGVFEGV-GNI-----ANFELDQYPLPGQTKYTIASKDYP 180
DB 802 DYTGYVTCHNSNLSEKALNSFNPTNLRGNVNLNENASFTLGKANLFG-----TIQSIGTS 856
QY 181 EVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRVNPQFAVPKGTDALVRVDFEHGNA 240
DB 857 QVN-----LKNSHWHLTGNSNVNQLNLTNGHILHNAQNDANKVTTNTLTNLSLNG 910
QY 241 -----YLENNYKVGEEKLPI-----PKLNOQT-----TRTAGNKIPVTM 275
DB 911 SFYVWVDFTNKSNKVVVKNKSATGNFTLQVADKGTGEPNHNLTLFDASNATRNNEVT-L 969
QY 276 ANAYLDN-----QSTYIVVEPILEKENOT-----DKPSILPOQKRN 311
DB 970 ANGSVDRGAWKYKLRNVNGRYDLNPEVEKRNQIVDTNITTPNDIQADAPS----- 1021

TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1098 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-923-992A-8

Query Match 5.1%; Score 141; DB 4; Length 1098;
 Best Local Similarity 20.0%; Pred. No. 0.0055;
 Matches 126; Conservative 95; Mismatches 207; Indels 202; Gaps 32;

```

QY 2 KOLDKKIEKI-----AGIMK--QYGVKRESIVVNKKNALIIYPHGDHHDPI--- 48
DB 203 EDAEVKREELKLFSSKAGLDQEIQEHVKET---SSEENT---QKVDEHYANSLQNL 256
QY 49 -----DEHKPYGIGHSHSNYELFKPEEGVAKKGNKVY--TGELTNVNVLL 93
DB 257 AOKSLELDKATTNEQATQVKNQFLENAQKLEIQPLIKETNVKLYKAMSESLQVEKEL 316
QY 94 KNSTFNQNTLANQKRVSEFPPELEKLGINMLVKLITPDGKVLKSGKVFGEVGV 153
DB 317 KHNEANLEDLVAKSEIVR-----EYEGKL--NQSKNL--PELKOLEEAAHSLKQVVE 367
QY 154 NI-ANFELDQPLPGQTFKYTIASKD-----YPE--VSYDG---TFTVPTSLAYK 197
DB 368 DFRKFKTSEQVTPKRLKRDLAANNENNOQKIELTVSPENITVYEGEDVKFTVTA---K 423
QY 198 MASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGKIKLPK 257
DB 424 SDSKT---TLDFSLLTKYNPSVS-----DRISTNKTNT---DNHKAETITKNLK 469
QY 258 LNOGTTRT-----AGNKIPVTFMANAYLDNQSTYIVVEPILKENDQDKPSILPOFKRN 311
DB 470 LNESQTVTLKAKDSSGNVVEKTF-----TITVQKKEK-----QVPKT 507
QY 312 KAQENSKLDEKV-EPEKTSKVEKEKLSGTGN-----STNSTLEEVPV----- 355
DB 508 PEQKDSKTEKVPQEPKSNKQNLQELIKSAQOELEKLEKAIKELMEQPEIPSNPEYGIQ 567
QY 356 -----DPVQEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSSGEVIK 395
DB 568 KSIWESQKEPIQEAITSFKKIIGDSSSKYTYEHYFNKYKSHFMNYQLHAQM-----EMLT 622
QY 396 KKNADFTGEAPOGNGENKPSNGKYST-----GTVENQPT-----NK----- 433
DB 623 RKVVOYMNKYPDNAEIKKTFESDMKRTKEDNYSLENDALKGYEKFYFLTPFNKIQIVD 682
QY 434 PADSLPEAPNEKVPKPNSTNGMLNPSGNVSGSDPMLDPALEEA-----PAVDPVQ 484
DB 683 DFDKKVEQDQAPPI-PENSE-----MDQAKERAKIAVSKYMSKVLGDGVH 725
QY 485 EKLEKFTASVGLGL-----DSVFENMD 506
DB 726 QHLQKNKNSKIVDLKFLEAIKQQTIFDID 755

```

RESULT 9
 US-08-923-992A-2
 Sequence 2, Application US/08923992A
 Patent No. 6280738
 GENERAL INFORMATION:
 APPLICANT: Tai, Joseph Y.
 APPLICANT: Blake, Milan S.
 TITLE OF INVENTION: No. 6280738-IgA Fc-Binding Forms of the Group B
 NUMBER OF INVENTIONS: Streptococcal Beta Antigens
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 STREET: 1100 New York Avenue, N.W., Suite 600
 CITY: Washington
 STATE: D.C.

COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentLIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/923,992A
 FILING DATE: 05-SEP-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/024,707
 FILING DATE: 06-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REGISTRATION NUMBER: 32,893
 REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1164 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-923-992A-2

Query Match 5.1%; Score 141; DB 4; Length 1164;
 Best Local Similarity 20.0%; Pred. No. 0.006;
 Matches 126; Conservative 95; Mismatches 207; Indels 202; Gaps 32;

```

QY 2 KOLDKKIEKI-----AGIMK--QYGVKRESIVVNKKNALIIYPHGDHHDPI--- 48
DB 240 EDAEVKREELKLFSSKAGLDQEIQEHVKET---SSEENT---QKVDEHYANSLQNL 293
QY 49 -----DEHKPYGIGHSHSNYELFKPEEGVAKKGNKVY--TGELTNVNVLL 93
DB 294 AOKSLELDKATTNEQATQVKNQFLENAQKLEIQPLIKETNVKLYKAMSESLQVEKEL 353
QY 94 KNSTFNQNTLANQKRVSEFPPELEKLGINMLVKLITPDGKVLKSGKVFGEVGV 153
DB 354 KHNEANLEDLVAKSEIVR-----EYEGKL--NQSKNL--PELKOLEEAAHSLKQVVE 404
QY 154 NI-ANFELDQPLPGQTFKYTIASKD-----YPE--VSYDG---TFTVPTSLAYK 197
DB 405 DFRKFKTSEQVTPKRVKRDLAANNENNOQKIELTVSPENITVYEGEDVKFTVTA---K 460
QY 198 MASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNYKVGKIKLPK 257
DB 461 SDSKT---TLDFSLLTKYNPSVS-----DRISTNKTNT---DNHKAETITKNLK 506
QY 258 LNOGTTRT-----AGNKIPVTFMANAYLDNQSTYIVVEPILKENDQDKPSILPOFKRN 311
DB 507 LNESQTVTLKAKDSSGNVVEKTF-----TITVQKKEK-----QVPKT 544
QY 312 KAQENSKLDEKV-EPEKTSKVEKEKLSGTGN-----STNSTLEEVPV----- 355
DB 545 PEQKDSKTEKVPQEPKSNKQNLQELIKSAQOELEKLEKAIKELMEQPEIPSNPEYGIQ 604
QY 356 -----DPVQEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSSGEVIK 395
DB 605 KSIWESQKEPIQEAITSFKKIIGDSSSKYTYEHYFNKYKSHFMNYQLHAQM-----EMLT 559
QY 396 KKNADFTGEAPOGNGENKPSNGKYST-----GTVENQPT-----NK----- 433
DB 660 RKVVOYMNKYPDNAEIKKTFESDMKRTKEDNYSLENDALKGYEKFYFLTPFNKIQIVD 719
QY 434 PADSLPEAPNEKVPKPNSTNGMLNPSGNVSGSDPMLDPALEEA-----PAVDPVQ 484
DB 720 DLDDKVEQDQAPPI-PENSE-----MDQAKERAKIAVSKYMSKVLGDGVH 762

```

QY 485 EKLEKFTASYGLGL-----DSVIFNMD 506
Db 763 OHLOKKNNSKIVDLFKELEAIKQOTIFDID 792

RESULT 10
US-08-923-992A-6
; Sequence 6, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1128 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-6

Query Match 5.1%; Score 140; DB 4; Length 1128;
Best Local Similarity 19.8%; Pred. No. 0.007;
Matches 125; Conservative 96; Mismatches 207; Indels 202; Gaps 32;

QY 2 KLDLKKIEKI-----AGIMKQY--GVKRESIVNVNKKNAIYPHGDHHDPI---48
Db 204 EDAEVKVRRELGLFSTRAGLQDEIHEVVKET---SSEENT---QKVDEHYANSLQNL 257

QY 49 -----DEHKPVGIGHSHENYELFKPEGVAKKEGKVI--TGEELTNVNNLL 93
Db 258 AQSLEELDKATTNEQATQYKNQFLENAQKLKEMQPLIKETNVKLYKAMSESLEQVEKEL 317

QY 94 KNSTFNQNTTANGOKRVSFSPPELEKLGINMLVKLITPDGKLVLEKVGKVFGEVGV 153
Db 318 KHNSANLEDLVAKSKEIVR-----EYEGKL--NQSKNL--PELKQLEEEAHSKLQOVVE 368

QY 154 NI-ANFELDQYPLFGQTFKYTIASKD-----YPE--VSVDG---TFTVPTSLAYK 197
Db 369 DFRKKFKTSEQVTPKKVRKRDLAANNQKIELTVSPENITVIEGEDVKFTVTA----K 424

QY 198 MASQTIYPPHAGDTYLRVNPQAVPKGTDALVRFVDEFGHAYLENNYKVGKIKLPIPK 257
Db 425 SDSKNT---TLDFSDLLTKYNPVS-----DRISTNKINT---DNHKAIEDITIKNLK 470

QY 258 LNAQGTIRI-----AGNKIPVTTFMANAYLDNQSTVIVEVPILEKENQTDKPSILPQFRN 311
Db 471 LNESQTVTLKARDDSGNVYKTF-----TITVQKKEK-----QVEKT 508

QY 312 KAQENSKLDEKV-EBPKTSEKVEKEKLSETGN-----STNSTLEEYPTV-----355
Db 509 PEQKDSKTEEKVPQEPKSNQNDKNQLOELIKSAQOELEKLEKAIKELMEQPEIPSPNPEYGIQ 568

QY 356 -----DPVQEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSEGEVIR 395
Db 569 KSIMESQKEPIQEAITSFKKIIGDSSSKYVTEHYFNKYKSDFMNYQLHAQM-----EMLT 623

QY 396 KNMAQFTGEAPQNGENKPSNGKYST-----GTVENQPT-----NK-----433
Db 624 RAVVQYMKYPDIAEIKKIFESDMARTKEDNNGVSGLENDALKGYPEKYLETFPFNKIKQIVD 683

QY 434 PADSLPEAFNEKPVKPNSTNDGMLNPECNVGSDPMLDPALEEA-----PAVDVQV 484
Db 684 DLDDKKVEQDPAPI-PENSE-----MDQAKEKAKIAVSKYMSKVLDDGVH 726

QY 485 EKLEKFTASYGLGL-----DSVIFNMD 506
Db 727 OHLOKKNNSKIVDLFKELEAIKQOTIFDID 756

RESULT 11
US-08-923-992A-4
; Sequence 4, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1104 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-4

Query Match 5.1%; Score 139; DB 4; Length 1104;
Best Local Similarity 19.8%; Pred. No. 0.0082;
Matches 125; Conservative 95; Mismatches 208; Indels 202; Gaps 32;

```
QY 2 KDLDKKIEERI-----AGIMQV--GVKRESIVVNEKNAIITYPHGDHHADPI--- 48
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 EDAEVKREELKGLFSSTAGLDQOIQEHVKET---SSEENT---QKVDEHYANSLQNL 262
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 49 -----DEHKPVGIGHSHSNYELFPBEGVAKKEGKNKY--TGEELTNVWLL 93
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 263 AOKSLELDKATNEAQTQVKNQFLENACKLKEIOLIKETNVKYKAMSELEQVEXEL 322
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 94 KNSTFNQNFLLANGOKRYSFPPPLEKLGINMLVCLITPDGKVLKESGKVFCEGVG 153
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 KHUSEANLODLVAKSEIVR---EYEGKL--NQSKNL--PELQLEEEAHSKLVQVVE 373
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 154 NI-ANFELDOPYLPQQTFFKYTIASKD-----YPE--VSYDG---TFTVPTSLAYK 197
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 374 HFRKFKTSQVTPKRVKRDLAANENNOQKIETLVSPENITVIEGEDVKFTVTA---K 429
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 198 MASOTIFYPHAGDTYLRVNPQFAVPKGTDALRVFDEFHGNAYLENNYKVGELPIPK 257
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 SDSKT---TLDFSLLTKYNPSVS-----DRISTNYKTNT---DNHKIAETIKNLK 475
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 258 LNQGTTT-----AGNKIPVTIFMANAYLDNQSTYIIVEPILEKENQTDKPSILPQFRN 311
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 476 LNSQVFTLAKDDSGNVKEKT-----TITVQKKEK-----QVPKKT 513
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 312 KAQENSKLDEKV--EKPTESEKVEKELSETGN-----STSNSTLEEVPV----- 355
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 514 PEQKHSKTEQNVQPEPKSNDKNGLOELIKSAQOELEKKAKEKMEQPEIPSPNPIYQ 573
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 356 -----DQVQEKVAKF-----ABSYGMKLENNFNMDGTIELYLPSEGVK 395
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 574 KSIWESQKEIQEAITSFNKIIIGDSSSKYTYEHYFNKYKSHFMYOLHAQM-----EILT 628
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 396 KNAADTGEAPQNGENKPSNGKYST-----GTVENQTE-----NK----- 433
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 629 RKVQVQVNNKYPDNAEIKKIFESDMKRTKEDNYGSLDALKGYFEKYFLTPTFNKIQIVD 688
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 434 PADSLPEAPNEKPVKPNSTDNMGNLPEGNVSGDPMLDPALEEA-----PAVDPVQ 484
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 689 DLQKVEQDOPAI--PENSE-----MDQAKEKAKIAVSKYMSKVLGDGVH 731
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 485 EKLEKFTASYGLGL-----DSVIFNMD 506
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 732 QHLOKKNHSHKIVDLFKLEAIKQOTIFDID 761
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-08-242-932-2
: Sequence 2, Application US/08242932
: Patent No. 5595740
: GENERAL INFORMATION:
: APPLICANT: Brady, L. Jeannine
: TITLE OF INVENTION: Cloning of No. 5595740-IgA Fc Binding Forms of
: TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: Saliwanchik & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/242,932
: FILING DATE: 16-MAY-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
```

```
REFERENCE/DOCKET NUMBER: UFI142
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 984 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-242-932-2

Query Match 5.08; Score 138.5; DB 1; Length 984;
Best Local Similarity 17.08; Pred. No. 0.0076;
Matches 101; Conservative 95; Mismatches 176; Indels 221; Gaps 25;

QY 1 MKDLDKKIEERIAGIMQYGVKRESIVVNEKNAIITYPHGD--HHADPIDEHKPVGI--- 56
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 LSNIDKELNHQKSQVEK---MAEQKGITNEDKDSMLKKIEDIRKQAOQADKEDAEVKVQ 210
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 57 --GSHSNYELFPBEGVAKKEGKNKYVTGELTNVWLLKNSITFNQNFLLANGOKRVSF 114
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 LEEEAHSLK-----KQVDEFKFKTSQVTPKRVKRDLAANE-----NNQKIEL 258
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 SFPPLEKLGINMLVCLITPDGKVLKESGKVEGEGVNIANFELDQYLPQGTFFYTI 174
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 TVSP-----NITVYE-----GEDVKFTV 277
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 ASKQYPSVSYDQTTVPTPSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALRVFD 234
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 278 TAKDSKTTLD-----FSDLLTKYNPSVS-----DRIST 306
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 EFHGNAYLENNYKVGELPIPKLNQGTTRT-----AGNKIPVTIFMANAYLDNQSTYIV 288
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 NYKTNT---DNHKIAETIKNLKINESQTVTLKAKDDSGNVKEKT----- 349
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 289 EYPILEKENQTDKPSILPQFRNKAQENSKLDEKV--EKPTESEKVEKELSETGN----- 342
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 TITVQKKEK-----QVPEKTEQKDSKTEKVPQPEPKSNDKNGLOELIKSAQOELEK 401
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 -STSNSTLEEVPV-----DPVQEKVAKF-----AESYGMK 372
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 402 LEKAIKELMEQPEIPSPNPEYGIQKSIWESQKEIQEAITSFKKIIIGDSSSKYTYEHFNK 461
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 373 LENVLNFMMDGTIELYLPSEGVKIKKNMADTGEAPQNGENKPSNGKYST-----GTVEN 427
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 YKSDFMNYQLHAQM-----EMLTRKVQVQVNNKYPDNAEIKKIFESDMKRTKEDNYGSLN 516
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 428 QTE-----NK-----PADSLPEAPNEKPVKPNSTDNMGNLPEGNVSGDPML 470
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 517 DALQGYFEKYFLTPTFNKIQIVDDLDKVKVEQDQPAI--PENSE-----M 559
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 471 DPALPEA-----PAVDPVQEKLEKFTASYGLGL-----DSVIFNMD 506
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 560 DQAKEKAKIAVSKYMSKVLGDGVHQLQKKNKSKIVDLFKLEAIKQOTIFDID 612
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-08-714-481-2
: Sequence 2, Application US/08714481
: Patent No. 5766606
: GENERAL INFORMATION:
: APPLICANT: Brady, L. Jeannine
: TITLE OF INVENTION: Cloning of No. 5766606-IgA Fc Binding Forms of
: TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESS: Saliwanchik & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,481
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF142
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-714-481-2

Query Match 5.0%; Score 138.5; DB 1; Length 984;
Best Local Similarity 17.0%; Pred. No. 0.0076;
Matches 101; Conservative 95; Mismatches 176; Indels 221; Gaps 25;

QY 1 MKDLDDKKEIKIAGIMKQGVKRESIVVNEKKNAILIYPHGD-HHHADPIDEHKPVGI--- 56
Db 154 LSNIDKELNHQKSOVEK---MAEQKGINEDKDSMLKIEDIRKQAQADKEDAEVRVQ 210
QY 57 --GHSNVELFKPEGVAKKGNKVVYGEELTNVNNLLKNSTFNQNTLANGOKRVSF 114
Db 211 LEEEAHSLK-----KQVDEFRKKFKTSEQVTPKKRVKRDLAANE-----NNQKI 258
QY 115 SPPELEKLGINMLVKLITPDGKVKESGVGEGVGNIANFELDQPLPGQTFKYTI 174
Db 259 TVSPE-----NITVE-----GSDVVKFTV 277
QY 175 ASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFVAPKGTDLVRVFD 234
Db 278 TAKSDSKTTLD-----FSDLLTKYNPSVS-----DRIST 306
QY 235 EFHGNAYLENNYKVGKIKPIPKLNQGTTRT-----AGNKIPVTFMANAYLDNQSTIV 288
Db 307 NYKNT---DNHKAETITKNLKNESQTVTLKAKDDSGNVVKTFF----- 349
QY 289 EYPILEKENQTDKPSILPOFKENKAQENSKLDEKV-EEPKTSEKVEKEKLSGTGN----- 342
Db 350 TITVQKKEK-----QVPKTPQKDSKTEEKVPOEPKSNKDKLOELIKSAQOELEK 401
QY 343 -STSNSTLSEVTV-----DPQVEKVAKF-----AESYGMK 372
Db 402 LEKAIKELMEQPEIPSPNPEYGIQKSIWESQKPEIQEATISFKKIIGDSSSKSYTFHYFNK 461
QY 373 LENVLNFMGDTIELYLPSPGEIKKNMADFTGAPQNGENKPSNGKYST-----GTVEN 427
Db 462 YKSDFNVLQHAQM-----EMLTRVQVYMKYKYPDIAEIKKIFESDMKRTKEDNVGSGLEN 516
QY 428 QPTE-----NK-----PADSLPEAPNEKPKVPENSTDNGLNPEGNVGSDEML 470
Db 517 DALIKGYFEKYLTPFNKIKQIVDDLDKKVQEQDQAPPI-PENSE-----M 559
QY 471 DPALPEA-----PAVDPPQKLEKFTASYGLGL-----DSVIFNMD 506
Db 560 DOAKERAKTAVSKYSKVLGDVGHQHLQKNNSQIVDLFKLEAIKQQTIFDID 612

RESULT 14

```

```

PCT-US95-061111-2
; Sequence 2, Application PC/TUS9506111
; GENERAL INFORMATION:
; APPLICANT: Street address: 186 Grinter Hall
; APPLICANT: City: Gainesville
; APPLICANT: State/Province: Florida
; APPLICANT: Country: US
; APPLICANT: Postal code/Zip: 32611
; APPLICANT: Phone number: 904-392-8929
; APPLICANT: Fax number: 904-392-6600
; APPLICANT: Telex number:
; TITLE OF INVENTION: Cloning of Non-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06111
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-061111-2

Query Match 5.0%; Score 138.5; DB 5; Length 984;
Best Local Similarity 17.0%; Pred. No. 0.0076;
Matches 101; Conservative 95; Mismatches 176; Indels 221; Gaps 25;

QY 1 MKDLDDKKEIKIAGIMKQGVKRESIVVNEKKNAILIYPHGD-HHHADPIDEHKPVGI--- 56
Db 154 LSNIDKELNHQKSOVEK---MAEQKGINEDKDSMLKIEDIRKQAQADKEDAEVRVQ 210
QY 57 --GHSNVELFKPEGVAKKGNKVVYGEELTNVNNLLKNSTFNQNTLANGOKRVSF 114
Db 211 LEEEAHSLK-----KQVDEFRKKFKTSEQVTPKKRVKRDLAANE-----NNQKI 258
QY 115 SPPELEKLGINMLVKLITPDGKVKESGVGEGVGNIANFELDQPLPGQTFKYTI 174
Db 259 TVSPE-----NITVE-----GSDVVKFTV 277
QY 175 ASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFVAPKGTDLVRVFD 234
Db 278 TAKSDSKTTLD-----FSDLLTKYNPSVS-----DRIST 306
QY 235 EFHGNAYLENNYKVGKIKPIPKLNQGTTRT-----AGNKIPVTFMANAYLDNQSTIV 288
Db 307 NYKNT---DNHKAETITKNLKNESQTVTLKAKDDSGNVVKTFF----- 349

```

```
QY 289 EVPILEKXNQTKPSILPQPKRNKAQENSKLDEKV-EPEKTSKVEKEKLESETGN----- 342
DB 350 TITVQKREK-----QVFKTPQKDSKTEBKVPQEPKSNKQLOELIKSAQOELEK 401
QY 343 -STNSSTLEEVPTV-----DPVQKVKAKF-----AESYGMK 372
DB 402 LEKAIKELMEQPEIPSPNEXYOKSIWESKKEPIQRAITSFKLIIGDSSSKYTYEHYFNK 461
QY 373 LENVLFNMDGTIELYLPDSGEVKKKNMADFTGEAPQNGENKPKSPENKSVST-----GTVEN 427
DB 462 YKSDFMNYQLHAQM-----EMLTRKVVQYNNKYPDMAETKKIFESDMKRTKEDNYGSLFN 516
QY 428 QPTE-----NK-----PADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVGSDPML 470
DB 517 DALGYPEKFTLTFNFIKOIIVDDLDKKVQDQDPAPI-PENSE-----M 559
QY 471 DPALEEA-----PAVDPVQOEKLEKFTASYGLGL-----DSVIFNMD 506
DB 560 DOAKEKAKIAVSKYMSKVLGDGVHQHLQKKNKSKIVDLFKLEAIKOOTIFDIID 612

RESULT 15
US-08-589-756-1
: Sequence 1, Application US/08589756
: Patent No. 5846547
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: STREPTOCOCCAL C5a PEPTIDASE VACCINE
: NUMBER OF SEQUENCES: 5
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/589,756
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-589-756-1

Query Match 4.88; Score 132.5; DB 2; Length 1164;
Best Local Similarity 23.48; Pred. No. 0.03;
Matches 91; Conservative 43; Mismatches 144; Indels 111; Gaps 19;

QY 17 KOYGVKRESIVVNKKNALIVPHGDHHDPIDEHKPVGIGHSHSNYELFKPEGVAKKE 76
DB 805 RHYVIHRHA---NGKPYAAISPNGDGNR-DYVQGF-----GTFLRNAKNLVAE--VLDKE 853
QY 77 GNKYVTGTELTNNVNLKN---STFNQNF-----TLANGOKRVSFSPFP-- 118
DB 854 GNVVMTSEVTEQVYNNVNDLASTLGTREISRWGDKDKAKVWANGTYTYRVRTPIIS 913
QY 119 --ELEKLGINMLVKLITPBGVLEKVSQVGFGEVGNIANF-----ELDQP 163
DB 914 SGAKEQHTDQDVIVDNTPE-----VATSATFSTEDRRUTLASKPOTSOP 958
QY 164 YLPQQTFKYITASKDYPEVSY-----DGTETVP-----TSLAYKMASQIFYPFHAG 210
DB 959 -VYRERIATYTWDELDLFTTYISPNEDGTFTLPEAEFTMEGATVPLKNSDFTYVVEDMAG 1017
QY 211 D-TYLRVNPQFAVPKGDALVRVDFEPHGNALYNNYKVGEEKLPIPKLNQGTTRTAGNK 269
DB 1018 NITVTPV-----TKLL-----EGHSNKPEDGSDQADPKPKETKPEQDGSQADPK 1063
QY 270 IPVTFMANYLDNSTIVIEVPILEKENQT--DKPSILPQFKENKAQENSKLDEKVEEPK 327
DB 1064 KPET-----KPGQDGGSGQTPDKKPKETKPE-KDSSGQTPGKTFQKGQPSR 1106
```

```
QY 328 TSEKVEKEKLESETGNSTNSSTLEEVPTVD 356
DB 1107 TLEK-----RSKRALATKASTROOLPTTN 1131

Search completed: May 13, 2003, 13:57:52
Job time : 15.1843 secs
```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: May 13, 2003, 13:53:57 ; Search time 13,0507 Seconds
(without alignments)
3723.142 Million cell updates/sec

Title: US-09-471-255-10
Perfect score: 2746
Sequence: 1 MKDLKKIEKIKAGIMKOY.....IELRLPSGEVKKNLSDPIA 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues
Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2746	100.0	569	9	US-09-884-465A-235 Sequence 235, App
2	2746	100.0	569	9	US-09-884-465A-255 Sequence 255, App
3	2746	100.0	627	9	US-09-884-465A-366 Sequence 366, App
4	2746	100.0	633	9	US-09-884-465A-350 Sequence 350, App
5	2746	100.0	633	9	US-09-884-465A-351 Sequence 351, App
6	2746	100.0	633	9	US-09-884-465A-362 Sequence 362, App
7	2746	100.0	894	9	US-09-884-465A-337 Sequence 337, App
8	2746	100.0	895	9	US-09-884-465A-347 Sequence 347, App
9	2746	100.0	900	9	US-09-884-465A-335 Sequence 335, App
10	2746	100.0	900	9	US-09-884-465A-339 Sequence 339, App
11	2746	100.0	900	9	US-09-884-465A-341 Sequence 341, App
12	2746	100.0	901	9	US-09-884-465A-343 Sequence 343, App
13	2746	100.0	901	9	US-09-884-465A-345 Sequence 345, App
14	2746	100.0	913	9	US-09-884-465A-384 Sequence 384, App
15	2746	100.0	999	9	US-09-884-465A-376 Sequence 376, App
16	2746	100.0	999	9	US-09-884-465A-377 Sequence 377, App
17	2746	100.0	1039	9	US-09-884-465A-6 Sequence 6, Appl
18	2746	100.0	1039	9	US-09-884-465A-6 Sequence 6, Appl
19	2746	100.0	1126	9	US-09-884-465A-383 Sequence 383, App

ALIGNMENTS

RESULT 1

US-09-884-465A-235
Sequence 235, Application US/09884465A
Publication No. US2003007293A1

GENERAL INFORMATION:

- APPLICANT: Shire Biochem, Inc.
- APPLICANT: Hamel, Josee
- APPLICANT: Brodeur, Bernard
- APPLICANT: Martin, Denis
- APPLICANT: Charland, Nathalie
- APPLICANT: Ouellet, Catherine
- TITLE OF INVENTION: Streptococcus Antigens
- FILE REFERENCE: 055190-0044
- CURRENT APPLICATION NUMBER: US/09/884,465A
- CURRENT FILING DATE: 2001-06-20
- PRIOR APPLICATION NUMBER: 60/212,683
- PRIOR FILING DATE: 2000-06-20
- NUMBER OF SEQ ID NOS: 384
- SOFTWARE: PatentIn version 3.1
- SEQ ID NO 235
- LENGTH: 569
- TYPE: PRT
- ORGANISM: Artificial Sequence
- FEATURE:
- OTHER INFORMATION: Unknown Organism

US-09-884-465A-235

Query Match 100.0%; Score 2746; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKDLKKIEKIKAGIMKOYGVKRESIVVNKEKALITYPHGDHHDHPIDHKVGVGHS	60
Db	42	MKDLKKIEKIKAGIMKOYGVKRESIVVNKEKALITYPHGDHHDHPIDHKVGVGHS	101
Qy	61	SNYELFKPEESGVAKCGNKVYTGEEITNVVNLKNSFTNNQNTLANGOKRVSFSPPEL	120
Db	102	SNYELFKPEESGVAKCGNKVYTGEEITNVVNLKNSFTNNQNTLANGOKRVSFSPPEL	161
Qy	121	EKLGINLWKLITPDGKVLKESGVKGVCNINAFELDPQYLPQGTFKYTIASKDYP	180
Db	162	EKLGINLWKLITPDGKVLKESGVKGVCNINAFELDPQYLPQGTFKYTIASKDYP	221

```
QY 181 EVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 240
DB 222 EVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 281
QY 241 YLENNYKVGKELPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTD 300
DB 282 YLENNYKVGKELPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTD 341
QY 301 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLNSTLEEVPTVDPVQE 360
DB 342 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLNSTLEEVPTVDPVQE 401
QY 361 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLNSTLEEVPTVDPVQE 420
DB 402 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLNSTLEEVPTVDPVQE 461
QY 421 STGTVENOPTENKPADSLPEAPNEKPVKPEKSTNMGMLNPEGNGSDPMLDPALEAPAV 480
DB 462 STGTVENOPTENKPADSLPEAPNEKPVKPEKSTNMGMLNPEGNGSDPMLDPALEAPAV 521
QY 481 DPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 528
DB 522 DPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 569

RESULT 2
US-09-884-465A-255
: Sequence 255, Application US/09884465A
: Publication No. US2003007293A1
: GENERAL INFORMATION:
: APPLICANT: Shire Biochem, Inc.
: APPLICANT: Hamel, Josee
: APPLICANT: Brodeur, Bernard
: APPLICANT: Martin, Denis
: APPLICANT: Charland, Nathalie
: APPLICANT: Ouellet, Catherine
: TITLE OF INVENTION: Streptococcus Antigens
: FILE REFERENCE: 055190-0044
: CURRENT APPLICATION NUMBER: US/09/884,465A
: CURRENT FILING DATE: 2001-06-20
: PRIOR APPLICATION NUMBER: 60/212,683
: PRIOR FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 384
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 255
: LENGTH: 569
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Unknown Organism
US-09-884-465A-255

Query Match 100.0%; Score 2746; DB 9; Length 569;
Best Local Similarity 100.0%; Pred. No. 1.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 MKDLCKIEEKIAGIMKQYGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSH 60
DB 42 MKDLCKIEEKIAGIMKQYGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSH 101
QY 61 SNYELFKPEEGVAKKGNKYVTGEEELTNVNLKNSFTNNQNTFLANGOKRVSFSPPEL 120
DB 102 SNYELFKPEEGVAKKGNKYVTGEEELTNVNLKNSFTNNQNTFLANGOKRVSFSPPEL 161
QY 121 EKKLGINMLVKLITPDGKLVLEKSGKVGEGVGNIANFELDQPLPGQTKFYTIASKDYP 180
DB 162 EKKLGINMLVKLITPDGKLVLEKSGKVGEGVGNIANFELDQPLPGQTKFYTIASKDYP 221
QY 181 EVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 240
DB 222 EVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 281
```

```
QY 241 YLENNYKVGKELPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTD 300
DB 282 YLENNYKVGKELPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTD 341
QY 301 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLNSTLEEVPTVDPVQE 360
DB 342 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLNSTLEEVPTVDPVQE 401
QY 361 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLNSTLEEVPTVDPVQE 420
DB 402 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLNSTLEEVPTVDPVQE 461
QY 421 STGTVENOPTENKPADSLPEAPNEKPVKPEKSTNMGMLNPEGNGSDPMLDPALEAPAV 480
DB 462 STGTVENOPTENKPADSLPEAPNEKPVKPEKSTNMGMLNPEGNGSDPMLDPALEAPAV 521
QY 481 DPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 528
DB 522 DPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 569

RESULT 3
US-09-884-465A-366
: Sequence 366, Application US/09884465A
: Publication No. US2003007293A1
: GENERAL INFORMATION:
: APPLICANT: Shire Biochem, Inc.
: APPLICANT: Hamel, Josee
: APPLICANT: Brodeur, Bernard
: APPLICANT: Martin, Denis
: APPLICANT: Charland, Nathalie
: APPLICANT: Ouellet, Catherine
: TITLE OF INVENTION: Streptococcus Antigens
: FILE REFERENCE: 055190-0044
: CURRENT APPLICATION NUMBER: US/09/884,465A
: CURRENT FILING DATE: 2001-06-20
: PRIOR APPLICATION NUMBER: 60/212,683
: PRIOR FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 384
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 366
: LENGTH: 627
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Unknown Organism
US-09-884-465A-366

Query Match 100.0%; Score 2746; DB 9; Length 627;
Best Local Similarity 100.0%; Pred. No. 1.4e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 MKDLCKIEEKIAGIMKQYGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSH 60
DB 100 MKDLCKIEEKIAGIMKQYGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSH 159
QY 61 SNYELFKPEEGVAKKGNKYVTGEEELTNVNLKNSFTNNQNTFLANGOKRVSFSPPEL 120
DB 160 SNYELFKPEEGVAKKGNKYVTGEEELTNVNLKNSFTNNQNTFLANGOKRVSFSPPEL 219
QY 121 EKKLGINMLVKLITPDGKLVLEKSGKVGEGVGNIANFELDQPLPGQTKFYTIASKDYP 180
DB 220 EKKLGINMLVKLITPDGKLVLEKSGKVGEGVGNIANFELDQPLPGQTKFYTIASKDYP 279
QY 181 EVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 240
DB 280 EVSYDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGNA 339
QY 241 YLENNYKVGKELPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTD 300
DB 340 YLENNYKVGKELPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENQTD 399
QY 301 KPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLNSTLEEVPTVDPVQE 360
```


Db	400	KPSILPQFKRNKAQENSKLDEKVEEPEKTSKVEKEKLSGTGNSNSTLEEVTVDPE	459
Qy	361	KVAFASYSYGMKLENLFFNMDGTIELYSPSGEVIKKNMADTGEAPQNGENKPSGVK	420
Db	460	KVAFASYSYGMKLENLFFNMDGTIELYSPSGEVIKKNMADTGEAPQNGENKPSGVK	519
Qy	421	STGTVENQOPTENKPADSLPEAPNEKPVKPSNSTDNGMLNPEGVGSDPMLDPALEAPAV	480
Db	520	STGTVENQOPTENKPADSLPEAPNEKPVKPSNSTDNGMLNPEGVGSDPMLDPALEAPAV	579
Qy	481	DPVQEKLEKTASYGLGDSVIFNMGGTIELRLPSPGEVIKKLSDFIA	528
Db	580	DPVQEKLEKTASYGLGDSVIFNMGGTIELRLPSPGEVIKKLSDFIA	627

```

RESULT 4
US-09-884-465A-350
; Sequence 350, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 350
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-350

```

[illegible]

```

DB 456 VKAFPAESYGMKLENLVFNMDGTIELVPSGEVTKKNMADPTGEAPGNGENKPSNGKY 525
QY 421 STGTVENOPTENKPADSLPEAFNEKPKVPKPNSTDNGLNPEGNVSGDPMLDPALEEAAPV 480
DB 536 STGTVENOPTENKPADSLPEAFNEKPKVPKPNSTDNGLNPEGNVSGDPMLDPALEEAAPV 585
QY 481 DPVQEKLEKFTASVGLGLDSVIFNWDGTIELRLPSGEVVKKNLSDFTA 528
DB 586 DPVQEKLEKFTASVGLGLDSVIFNWDGTIELRLPSGEVVKKNLSDFTA 633

RESULT 5
US-03-884-465A-351
; Sequence 351, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 351
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-03-884-465A-351

Query March 100.0%; Score 2746; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.4e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Query Match 100.0%; Score 2746; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.4e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKD	K	K	I	E	E	K	I	A	G	I	M	K	O	Y	G	V	K	R	E	S	I	V	V	N	E	K	N	A	I	I	P	H	G	D	H	H	A	D	I	D	E	K	P	V	G	I	G	H	S	60						
Db	106	MKD	K	K	I	E	E	K	I	A	G	I	M	K	O	Y	G	V	K	R	E	S	I	V	V	N	E	K	N	A	I	I	P	H	G	D	H	H	A	D	I	D	E	K	P	V	G	I	G	H	S	165						
Qy	61	S	N	E	L	F	P	E	E	G	V	A	K	E	G	N	K	Y	I	T	G	E	L	N	V	N	L	L	K	N	S	T	F	N	N	O	T	L	A	N	G	K	R	V	S	F	S	E	P	P	E	L	120					
Db	166	S	N	E	L	F	P	E	E	G	V	A	K	E	G	N	K	Y	I	T	G	E	L	N	V	N	L	L	K	N	S	T	F	N	N	O	T	L	A	N	G	K	R	V	S	F	S	E	P	P	E	L	225					
Qy	121	E	K	L	G	I	N	M	L	V	K	L	I	P	D	G	K	V	L	E	K	V	S	G	K	V	G	E	G	V	N	I	A	N	F	E	L	D	P	Y	L	P	Q	T	F	K	Y	T	I	A	S	K	D	Y	180			
Db	226	E	K	L	G	I	N	M	L	V	K	L	I	P	D	G	K	V	L	E	K	V	S	G	K	V	G	E	G	V	N	I	A	N	F	E	L	D	P	Y	L	P	Q	T	F	K	Y	T	I	A	S	K	D	Y	285			
Qy	181	E	V	S	D	G	T	F	T	P	T	S	L	A	Y	K	M	A	S	O	T	I	F	P	H	A	G	D	T	I	V	N	P	O	F	A	V	P	K	G	T	D	A	L	V	R	V	E	D	E	H	G	N	A	240			
Db	286	E	V	S	D	G	T	F	T	P	T	S	L	A	Y	K	M	A	S	O	T	I	F	P	H	A	G	D	T	I	V	N	P	O	F	A	V	P	K	G	T	D	A	L	V	R	V	E	D	E	H	G	N	A	345			
Qy	241	Y	L	E	N	N	Y	K	V	G	E	I	K	L	P	I	P	K	L	N	G	T	T	R	A	G	N	K	I	P	V	T	F	M	A	N	A	I	L	D	N	G	S	T	I	V	E	P	I	L	E	K	N	O	T	300		
Db	346	Y	L	E	N	N	Y	K	V	G	E	I	K	L	P	I	P	K	L	N	G	T	T	R	A	G	N	K	I	P	V	T	F	M	A	N	A	I	L	D	N	G	S	T	I	V	E	P	I	L	E	K	N	O	T	405		
Qy	301	K	P	S	T	I	P	O	F	K	R	N	K	A	G	E	S	K	L	D	E	K	V	E	E	P	K	T	S	E	K	V	E	K	E	L	S	E	T	G	N	S	T	S	N	S	T	L	E	E	V	T	P	D	V	O	E	360
Db	406	K	P	S	T	I	P	O	F	K	R	N	K	A	G	E	S	K	L	D	E	K	V	E	E	P	K	T	S	E	K	V	E	K	E	L	S	E	T	G	N	S	T	S	N	S	T	L	E	E	V	T	P	D	V	O	E	465
Qy	361	K	V	A	F	A	S	Y	G	M	K	L	E	N	V	L	F	N	D	G	T	I	E	L	Y	L	P	S	G	V	I	K	N	M	A	D	F	T	C	E	A	P	O	G	N	G	E	N	K	P	S	E	N	G	V	420		
Db	466	K	V	A	F	A	S	Y	G	M	K	L	E	N	V	L	F	N	D	G	T	I	E	L	Y	L	P	S	G	V	I	K	N	M	A	D	F	T	C	E	A	P	O	G	N	G	E	N	K	P								

QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFA 528
 Db 586 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFA 633

RESULT 6

US-09-884-465A-361
 ; Sequence 361, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 053190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 361
 ; LENGTH: 633
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Unknown Organism
 US-09-884-465A-361

Query Match 100.0%; Score 2746; DB 9; Length 633;
 Best Local Similarity 100.0%; Pred. No. 1.4e-179;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEKTAGIMKQYGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGIGSH 60
 Db 106 MKDLKKIEKTAGIMKQYGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGIGSH 165

QY 61 SNYELFKPEEGVAKKEGKNVYTGEELTNVVNLKNSFNQNTFLANGQKRVSFSPPEL 120
 Db 166 SNYELFKPEEGVAKKEGKNVYTGEELTNVVNLKNSFNQNTFLANGQKRVSFSPPEL 225

QY 121 EKKLGINMLVKLITPDGKVLKESKVGKGVGEGVGNIANFELDPYLPQGTFFKVTIASKDYP 180
 Db 226 EKKLGINMLVKLITPDGKVLKESKVGKGVGEGVGNIANFELDPYLPQGTFFKVTIASKDYP 285

QY 181 EVSYDGTFTVPTSLAYKMASQITFYPPHAGDTYLRVNPQFAVPKGTDALVRVDEFHGN 240
 Db 286 EVSYDGTFTVPTSLAYKMASQITFYPPHAGDTYLRVNPQFAVPKGTDALVRVDEFHGN 345

QY 241 YLENNYKVGEEKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTD 300
 Db 346 YLENNYKVGEEKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTD 405

QY 301 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTVDPVQE 360
 Db 406 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTVDPVQE 465

QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVKKKMAADFTGEAPQNGENKPSGENKV 420
 Db 466 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVKKKMAADFTGEAPQNGENKPSGENKV 525

QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 480
 Db 526 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 585

QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFA 528
 Db 586 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFA 633

RESULT 8
 US-09-884-465A-337
 ; Sequence 337, Application US/09884465A
 ; Publication No. US20030077293A1

RESULT 7
 US-09-884-465A-362
 ; Sequence 362, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 053190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 362
 ; LENGTH: 633
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Unknown Organism
 US-09-884-465A-362

Query Match 100.0%; Score 2746; DB 9; Length 633;
 Best Local Similarity 100.0%; Pred. No. 1.4e-179;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKDLKKIEKTAGIMKQYGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGIGSH 60
 Db 106 MKDLKKIEKTAGIMKQYGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGIGSH 165

QY 61 SNYELFKPEEGVAKKEGKNVYTGEELTNVVNLKNSFNQNTFLANGQKRVSFSPPEL 120
 Db 166 SNYELFKPEEGVAKKEGKNVYTGEELTNVVNLKNSFNQNTFLANGQKRVSFSPPEL 225

QY 121 EKKLGINMLVKLITPDGKVLKESKVGKGVGEGVGNIANFELDPYLPQGTFFKVTIASKDYP 180
 Db 226 EKKLGINMLVKLITPDGKVLKESKVGKGVGEGVGNIANFELDPYLPQGTFFKVTIASKDYP 285

QY 181 EVSYDGTFTVPTSLAYKMASQITFYPPHAGDTYLRVNPQFAVPKGTDALVRVDEFHGN 240
 Db 286 EVSYDGTFTVPTSLAYKMASQITFYPPHAGDTYLRVNPQFAVPKGTDALVRVDEFHGN 345

QY 241 YLENNYKVGEEKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTD 300
 Db 346 YLENNYKVGEEKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSTIYVEVPILEKENQTD 405

QY 301 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTVDPVQE 360
 Db 406 KPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTLEEVPTVDPVQE 465

QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVKKKMAADFTGEAPQNGENKPSGENKV 420
 Db 466 KVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVKKKMAADFTGEAPQNGENKPSGENKV 525

QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 480
 Db 526 STGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPEGNVGSDDPMLDPALEAPAV 585

QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFA 528
 Db 586 DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGEVKKKNSDFA 633

GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 337
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-337

Query Match 100.0%; Score 2746; DB 9; Length 894;
Best Local Similarity 100.0%; Pred. No. 2.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKDLKKIEKIAIGIMKQYGVKRESIVVNEKNAIYPHGDHHDADPIDEHKPVGICHSH 60
Db 367 MKDLKKIEKIAIGIMKQYGVKRESIVVNEKNAIYPHGDHHDADPIDEHKPVGICHSH 426
QY 61 SNYELFKPEGVAKKEGKQYVYTGELTNVYLLKNSFTNNQNTFLANGQKRVFSFPPEL 120
Db 427 SNYELFKPEGVAKKEGKQYVYTGELTNVYLLKNSFTNNQNTFLANGQKRVFSFPPEL 486
QY 121 EKKLGIMLVKLTIPDGKYLEKVGSGKVFGEVGNIANFELDQYPLPGQTFKYTIASKDYP 180
Db 487 EKKLGIMLVKLTIPDGKYLEKVGSGKVFGEVGNIANFELDQYPLPGQTFKYTIASKDYP 546
QY 181 EVSVDGFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 240
Db 547 EVSVDGFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 606
QY 241 YLENNYKVGEEKIPLKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTD 300
Db 607 YLENNYKVGEEKIPLKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTD 666
QY 301 KPSTLPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLTSETGNSNSTLEEVTVDVQOE 360
Db 667 KPSTLPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLTSETGNSNSTLEEVTVDVQOE 726
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIVKKNMADFTGEAPQNGENKPSGENKV 420
Db 727 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIVKKNMADFTGEAPQNGENKPSGENKV 786
QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAV 480
Db 787 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAV 846
QY 481 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIVKKNLSDFIA 528
Db 847 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIVKKNLSDFIA 894

RESULT 9
US-09-884-465A-347
; Sequence 347, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis

APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: Patentin version 3.1
SEQ ID NO 347
LENGTH: 895
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-347

Query Match 100.0%; Score 2746; DB 9; Length 895;
Best Local Similarity 100.0%; Pred. No. 2.3e-179;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKDLKKIEKIAIGIMKQYGVKRESIVVNEKNAIYPHGDHHDADPIDEHKPVGICHSH 60
Db 368 MKDLKKIEKIAIGIMKQYGVKRESIVVNEKNAIYPHGDHHDADPIDEHKPVGICHSH 427
QY 61 SNYELFKPEGVAKKEGKQYVYTGELTNVYLLKNSFTNNQNTFLANGQKRVFSFPPEL 120
Db 428 SNYELFKPEGVAKKEGKQYVYTGELTNVYLLKNSFTNNQNTFLANGQKRVFSFPPEL 487
QY 121 EKKLGIMLVKLTIPDGKYLEKVGSGKVFGEVGNIANFELDQYPLPGQTFKYTIASKDYP 180
Db 488 EKKLGIMLVKLTIPDGKYLEKVGSGKVFGEVGNIANFELDQYPLPGQTFKYTIASKDYP 547
QY 181 EVSVDGFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 240
Db 548 EVSVDGFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAVPGKTDALVRVDFEHGNA 607
QY 241 YLENNYKVGEEKIPLKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTD 300
Db 608 YLENNYKVGEEKIPLKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVVEVPILEKENQTD 667
QY 301 KPSTLPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLTSETGNSNSTLEEVTVDVQOE 360
Db 668 KPSTLPQFKRNKAQENSKLDEKVEEPTSEKVEKEKLTSETGNSNSTLEEVTVDVQOE 727
QY 361 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIVKKNMADFTGEAPQNGENKPSGENKV 420
Db 728 KVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIVKKNMADFTGEAPQNGENKPSGENKV 787
QY 421 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAV 480
Db 788 STGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVSDPMLDPALEAPAV 847
QY 481 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIVKKNLSDFIA 528
Db 848 DPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIVKKNLSDFIA 895

RESULT 10
US-09-884-465A-335
; Sequence 335, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A

QY 61 SNYELFKPEGVAKKEGKVVYTGELTNVNVLLKKNSTNNQNTLANGOKRVSFSPPEL 120
DB 434 SNYELFKPEGVAKKEGKVVYTGELTNVNVLLKKNSTNNQNTLANGOKRVSFSPPEL 493
QY 121 EKKLGINMLVKLITPDGKLVKSVKVGEGVGNITANFELDQPYLPQGTFFYTTASKDYP 180
DB 494 EKKLGINMLVKLITPDGKLVKSVKVGEGVGNITANFELDQPYLPQGTFFYTTASKDYP 553
QY 181 EYSYDGTFTVPTSLAYKVASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFHNA 240
DB 554 EYSYDGTFTVPTSLAYKVASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFHNA 613
QY 241 YLENNYKVGKIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKQTD 300
DB 614 YLENNYKVGKIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKQTD 673
QY 301 KPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSKSETGNSNSTLEEVPTVDPVOE 360
DB 674 KPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSKSETGNSNSTLEEVPTVDPVOE 733
QY 361 KVAFAESYCMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSSENGKV 420
DB 734 KVAFAESYCMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSSENGKV 793
QY 421 STGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPGVNSDPMLEAPAV 480
DB 794 STGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPGVNSDPMLEAPAV 853
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFIA 528
DB 854 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFIA 901

RESULT 15

US-09-884-465A-384
; Sequence 384, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 384
; LENGTH: 913
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC.FEATURE
; LOCATION: (1)-(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (344)..(344)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (345)..(345)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-384

Query Match

100.0%; Score 2746; DB 9; Length 913;

Best Local Similarity 100.0%; Pred. No. 2.3e-179;

Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLDKKLEEKIAGIMKQGVKRESIIVWNKEKNAIYPHGDHHDADIDEHKPVGIGHSH 60
DB 386 MKLDKKLEEKIAGIMKQGVKRESIIVWNKEKNAIYPHGDHHDADIDEHKPVGIGHSH 445
QY 61 SNYELFKPEGVAKKEGKVVYTGELTNVNVLLKKNSTNNQNTLANGOKRVSFSPPEL 120
DB 446 SNYELFKPEGVAKKEGKVVYTGELTNVNVLLKKNSTNNQNTLANGOKRVSFSPPEL 505
QY 121 EKKLGINMLVKLITPDGKLVKSVKVGEGVGNITANFELDQPYLPQGTFFYTTASKDYP 180
DB 506 EKKLGINMLVKLITPDGKLVKSVKVGEGVGNITANFELDQPYLPQGTFFYTTASKDYP 565
QY 181 EYSYDGTFTVPTSLAYKVASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFHNA 240
DB 566 EYSYDGTFTVPTSLAYKVASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFHNA 625
QY 241 YLENNYKVGKIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKQTD 300
DB 626 YLENNYKVGKIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILKEKQTD 685
QY 301 KPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSKSETGNSNSTLEEVPTVDPVOE 360
DB 686 KPSILPOFKRNKAQENSKLDEKVEEPKTSKVEKEKLSKSETGNSNSTLEEVPTVDPVOE 745
QY 361 KVAFAESYCMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSSENGKV 420
DB 746 KVAFAESYCMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQNGENKPSSENGKV 805
QY 421 STGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPGVNSDPMLEAPAV 480
DB 806 STGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPGVNSDPMLEAPAV 865
QY 481 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFIA 528
DB 866 DPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVKKNLSDFIA 913

Search completed: May 13, 2003, 13:58:53

Job time : 15.0507 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:49:42 ; Search time 35.8944. Seconds
(without alignments)
3118.326 Million cell updates/sec

Title: US-09-471-255-16

Perfect score: 4396

Sequence: 1 CAYALNHRSEKNNRVS.....GTIELRPSGVKKNLLIS 840

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4396	100.0	840	21	AA12751
2	4396	100.0	840	23	AAU76151
3	4276.5	97.3	1019	21	AA12722
4	4276.5	97.3	1019	21	AA12751
5	4276.5	97.3	1019	23	AAU84021
6	4276.5	97.3	1039	21	AA12715
7	4276.5	97.3	1039	23	AAU75932
8	4274.5	97.2	1019	21	AA12750
9	4268.5	97.1	1019	21	AA12748
10	4268.5	97.1	1019	21	AA12752

11	4268.5	97.1	1019	21	AA12753	Streptococcus pneu
12	4264.5	97.0	1019	21	AA12749	Streptococcus pneu
13	3402.5	77.4	807	23	AAU83997	Truncated variant
14	3402.5	77.4	1152	23	AAU84054	S. pneumoniae deri
15	3402.5	77.4	1238	23	AAU84056	S. pneumoniae deri
16	3402.5	77.4	1365	23	AAU84057	S. pneumoniae deri
17	3402.5	77.4	1378	23	AAU84053	S. pneumoniae deri
18	3241	73.7	644	23	AAU83996	Truncated variant
19	3237	73.6	907	23	AAU84051	S. pneumoniae deri
20	3228	73.4	907	23	AAU84055	S. pneumoniae deri
21	3219	73.2	907	23	AAU84053	S. pneumoniae deri
22	3215	73.1	901	23	AAU84067	S. pneumoniae deri
23	3202.5	72.9	632	23	AAU83998	Truncated variant
24	3200.5	72.8	902	23	AAU84076	S. pneumoniae deri
25	3193.5	72.6	632	23	AAU84002	Truncated variant
26	3193.5	72.6	632	23	AAU84006	Truncated variant
27	3193.5	72.6	632	23	AAU84013	Truncated variant
28	3193.5	72.6	632	23	AAU84014	Truncated variant
29	3191	72.6	901	23	AAU84088	S. pneumoniae deri
30	3184.5	72.4	632	23	AAU84003	Truncated variant
31	3184.5	72.4	632	23	AAU84004	Truncated variant
32	3184.5	72.4	632	23	AAU84009	Truncated variant
33	3184.5	72.4	632	23	AAU84012	Truncated variant
34	3184.5	72.4	901	23	AAU84056	S. pneumoniae deri
35	3184.5	72.4	907	23	AAU84080	S. pneumoniae deri
36	3184.5	72.4	907	23	AAU84084	S. pneumoniae deri
37	3183.5	72.4	632	23	AAU84015	Truncated variant
38	3183.5	72.4	632	23	AAU84016	Truncated variant
39	3182.5	72.4	902	23	AAU84078	S. pneumoniae deri
40	3182	72.4	901	23	AAU84072	S. pneumoniae deri
41	3182	72.4	901	23	AAU84074	S. pneumoniae deri
42	3178.5	72.3	895	23	AAU84080	S. pneumoniae deri
43	3178	72.3	895	23	AAU84070	S. pneumoniae deri
44	3174.5	72.2	632	23	AAU84007	Truncated variant
45	3174.5	72.2	632	23	AAU84017	Truncated variant

ALIGNMENTS

RESULT 1
AA12721
ID AA12721 standard; Protein: 840 AA.
XX
AC AA12721;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae SP63 BVH-3 protein antigen SEQ ID NO:16.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX DR N-PSDB; AAA65738.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
otitis media, bacteraemia and/or pneumonia -

```
XX PS Claim 18; Fig 19; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae SP63 BVH-3 protein antigen.
XX SQ Sequence 840 AA;

Query Match 100.0%; Score 4396; DB 21; Length 840;
Best Local Similarity 100.0%; Pred. No. 1.le-261;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAYALNQHRSQENKNNRYSYVDGSSQSKSENLTDPQVSQKREGIOAEQIVIKITDQGYV 60
DB 1 CAYALNQHRSQENKNNRYSYVDGSSQSKSENLTDPQVSQKREGIOAEQIVIKITDQGYV 60
QY 61 TSHGDHYHYNGKVPYDALFSEBLLMKDPNYQLKDADIVNEVKGYYIIKVDGKYVYVYKLD 120
DB 61 TSHGDHYHYNGKVPYDALFSEBLLMKDPNYQLKDADIVNEVKGYYIIKVDGKYVYVYKLD 120
QY 121 AAHADNVRTKDEINROKQEHVKONKSVNNAVARSGQRYTTNDGVFNPADIEDTGN 180
DB 121 AAHADNVRTKDEINROKQEHVKONKSVNNAVARSGQRYTTNDGVFNPADIEDTGN 180
QY 181 YIVPHGHHYHYIPKSDLSASELAHAGKAGNNQPSQSYSTPSPSLPINPGTSHEKH 240
DB 181 YIVPHGHHYHYIPKSDLSASELAHAGKAGNNQPSQSYSTPSPSLPINPGTSHEKH 240
QY 241 EEDGYGFDAKRIITAEDESGVMSGHBNHYFFKKDLTEQIKAAQKHLSEVKTSHNGLDS 300
DB 241 EEDGYGFDAKRIITAEDESGVMSGHBNHYFFKKDLTEQIKAAQKHLSEVKTSHNGLDS 300
QY 301 LSSHEQDYPNNAKMDLDRKIEEKLKAGIMQYGVKRESIVNNEKNATIIYPHGDHHHAD 360
DB 301 LSSHEQDYPNNAKMDLDRKIEEKLKAGIMQYGVKRESIVNNEKNATIIYPHGDHHHAD 360
QY 361 PIDEHKPVGTGHSHSNVYELFPEEGVAKKGNKYVTGEELTNVNLKSTFNQNFLLA 420
DB 361 PIDEHKPVGTGHSHSNVYELFPEEGVAKKGNKYVTGEELTNVNLKSTFNQNFLLA 420
QY 421 NGQKRVSFPPPELEKKGINMLVLTITPDGKLVLEKVGKVGEGVGNIANFELDQPYLP 480
DB 421 NGQKRVSFPPPELEKKGINMLVLTITPDGKLVLEKVGKVGEGVGNIANFELDQPYLP 480
QY 481 GQTFKYTIASKDPEVSYDGTFTVPTSLAYKMASQITFVFFHAGDTYLRVNPQFAYPKGT 540
DB 481 GQTFKYTIASKDPEVSYDGTFTVPTSLAYKMASQITFVFFHAGDTYLRVNPQFAYPKGT 540
QY 541 DALVRVDEFGHAYLNNYKVGEIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSY 600
DB 541 DALVRVDEFGHAYLNNYKVGEIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSY 600
QY 601 IVEVPILKENQDTPKPSILPOFKRNKAQENSKLDEKVEPKTSEKVEKELSTGNTSN 660
DB 601 IVEVPILKENQDTPKPSILPOFKRNKAQENSKLDEKVEPKTSEKVEKELSTGNTSN 660
QY 661 STLEEVPTVPDQVEKAKFAESYCKMLNVLFNMDGTIELYLSGEVYKKNMADTFGEAP 720
DB 661 STLEEVPTVPDQVEKAKFAESYCKMLNVLFNMDGTIELYLSGEVYKKNMADTFGEAP 720
QY 721 QGNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGNVGS 780
DB 721 QGNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGNVGS 780
QY 781 DPMILDSALEAPAVDPQVEKLEKFTASYGGLDSVIFNMDGTIELRLPSEGVYKKNLLIS 840
DB 781 DPMILDSALEAPAVDPQVEKLEKFTASYGGLDSVIFNMDGTIELRLPSEGVYKKNLLIS 840
```

```
DB 781 DPMILDSALEAPAVDPQVEKLEKFTASYGGLDSVIFNMDGTIELRLPSEGVYKKNLLIS 840

RESULT 2
AAU76151
ID AAU76151 standard; Protein: 840 AA.
XX AC AAU76151;
XX DT 08-MAY-2002 (first entry)
XX DE Streptococcus pneumoniae BVH-3 protein version #2.
XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX KW pneumonia; streptococcal bacterial infection.
XX OS Streptococcus pneumoniae.
XX PN WO200198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX PI Hamel J, Ouellet C, Charla C, Charland N, Martin D, Brodeur B;
XX WPI: 2002-122272/16.
XX DR N-PSDB: ABK15105.
XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia.
XX PS Example 1; Fig 10; 113pp; English.
XX CC The invention describes an isolated polypeptide (I) with 70-90%
XX CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX CC comprising (I) is useful for therapeutic or prophylactic treatment of
XX CC meningitis, otitis media, bacteraemia or pneumonia infection in an
XX CC individual susceptible to these disorders (II) is also useful for
XX CC therapeutic or prophylactic treatment of any streptococcal bacterial
XX CC infection (e.g., caused by Streptococcus pneumoniae, group A
XX CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
XX CC Staphylococcus aureus) in an individual susceptible to the infection.
XX CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX CC techniques. The Streptococcus polypeptides are useful in a diagnostic
XX CC test for S. pneumoniae infection. (III) is useful for designing DNA
XX CC probes for use in detecting the presence of Streptococcus in a biological
XX CC sample suspected of containing the bacteria. The DNA probes may also be
XX CC used for detecting circulating S. pneumonia nucleic acid in a sample for
XX CC diagnosing streptococcal infections. This is the amino acid sequence of
XX CC Streptococcus pneumoniae protein BVH-3, used to create the antigenic
XX CC peptides described in the method of the invention.
XX SQ Sequence 840 AA;

Query Match 100.0%; Score 4396; DB 23; Length 840;
Best Local Similarity 100.0%; Pred. No. 1.le-261;
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAYALNQHRSQENKNNRYSYVDGSSQSKSENLTDPQVSQKREGIOAEQIVIKITDQGYV 60
DB 1 CAYALNQHRSQENKNNRYSYVDGSSQSKSENLTDPQVSQKREGIOAEQIVIKITDQGYV 60
QY 61 TSHGDHYHYNGKVPYDALFSEBLLMKDPNYQLKDADIVNEVKGYYIIKVDGKYVYVYKLD 120
DB 61 TSHGDHYHYNGKVPYDALFSEBLLMKDPNYQLKDADIVNEVKGYYIIKVDGKYVYVYKLD 120
```



```
Db 61 TSHGDHYHYNGKVPYDALFESEELMKDPNYQLKDADIVREVKGYYIKVDGKYVYVLKD 120
Qy 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQGRYTTNDGYVFNPADIIDTGN 180
Db 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQGRYTTNDGYVFNPADIIDTGN 180
Qy 181 YIVPHGCHYHYIPKSDLSASASAAKHAHLAKGNMPSQLSYSTSPSLPINPOTSHEKH 240
Db 181 YIVPHGCHYHYIPKSDLSASASAAKHAHLAKGNMPSQLSYSTSPSLPINPOTSHEKH 240
Qy 241 EEDCYGFDANRIIAEDSGFVMSHGDHNYHFFKKDLTEEQIKAAQKHLVEVKTSHNGLDS 300
Db 241 EEDCYGFDANRIIAEDSGFVMSHGDHNYHFFKKDLTEEQIKAAQKHLVEVKTSHNGLDS 300
Qy 301 LSSHEQDYFPSNAKEMKDLKIEKIAIGIMKQYGVKRESIVVNNKRNIAIYPHGDHHDH 360
Db 301 LSSHEQDYFPSNAKEMKDLKIEKIAIGIMKQYGVKRESIVVNNKRNIAIYPHGDHHDH 360
Qy 361 PIDEHKPVGIGHSHSNVELKPEEGVAKKGNKYVTGEBELTNVNLKSTFNQNTFLA 420
Db 361 PIDEHKPVGIGHSHSNVELKPEEGVAKKGNKYVTGEBELTNVNLKSTFNQNTFLA 420
Qy 421 NGQRVSEFPPPELEKKLGINMLVKLITPDGKYLEKYSKGVFGEVGNIANFELDQYLP 480
Db 421 NGQRVSEFPPPELEKKLGINMLVKLITPDGKYLEKYSKGVFGEVGNIANFELDQYLP 480
Qy 481 GQTKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPEKT 540
Db 481 GQTKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPEKT 540
Qy 541 DALVRYDFEFGNAYLENNYKVGGEIKLPKPLNOGTRTAGNKIPVTFMANAYLDNOSTY 600
Db 541 DALVRYDFEFGNAYLENNYKVGGEIKLPKPLNOGTRTAGNKIPVTFMANAYLDNOSTY 600
Qy 601 IVEVPILEKENQTDKPSLPOFKRNKAQENSKLDEKVEEPEKTSKYEKELSETGNSTSN 660
Db 601 IVEVPILEKENQTDKPSLPOFKRNKAQENSKLDEKVEEPEKTSKYEKELSETGNSTSN 660
Qy 661 STLEEVPYDVPQEKVAKFAESYGNKLENVLFNMDGTIELYLPSEVYIKKNMADFTGEAP 720
Db 661 STLEEVPYDVPQEKVAKFAESYGNKLENVLFNMDGTIELYLPSEVYIKKNMADFTGEAP 720
Qy 721 QNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVYPENSTDNGLNPEGNVGS 780
Db 721 QNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVYPENSTDNGLNPEGNVGS 780
Qy 781 DPMILDSALEAPAVDPVQEKLEKFTASYGLGLDSVIENMDGTIELRLPSEVYIKKNLLIS 840
Db 781 DPMILDSALEAPAVDPVQEKLEKFTASYGLGLDSVIENMDGTIELRLPSEVYIKKNLLIS 840

RESULT 3
AAB12722
ID AAB12722 standard; Protein; 1019 AA.
XX
AC AAB12722;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO2000039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
```

```
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
XX WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
XX
XX Claim 18; Fig 20; 105pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-3M protein antigen.
XX
SQ Sequence 1019 AA:

Query Match 97.3%; Score 4276.5; DB 21; Length 1019;
Best Local Similarity 82.3%; Pred. No. 3.le-254;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

Qy 1 CAYALNQRSQENKNNRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGVY 60
Db 1 CAYALNQRSQENKNNRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGVY 60
Qy 61 TSHGDHYHYNGKVPYDALFESEELMKDPNYQLKDADIVNEKGGYIIKVDGKYVYVLKD 120
Db 61 TSHGDHYHYNGKVPYDALFESEELMKDPNYQLKDADIVNEKGGYIIKVDGKYVYVLKD 120
Qy 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQGRYTTNDGYVFNPADIIDTGN 180
Db 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQGRYTTNDGYVFNPADIIDTGN 180
Qy 181 YIVPHGCHYHYIPKSDLSASASAAKHAHLAKGNMPSQLSYSTSPSLPINPOTSHEKH 223
Db 181 YIVPHGCHYHYIPKSDLSASASAAKHAHLAKGNMPSQLSYSTSPSLPINPOTSHEKH 240
Qy 224 ----- 223
Db 241 PANKSENLOSLLKELXDSPSAQRYSSDGLVFPDAKIISRTPNGVAIPHGDHYHFI 300
Qy 224 ----- 223
Db 301 LSALEBKIAKRWVPISGTGSTVSTNAKPNEVWVSSLSNPPSLTTSKELSSASDCYIFN 360
Qy 224 -----TPSPSLPINPGTSHKHEED 243
Db 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIQGPTLPNNSLATPSPLPINPGTSHKHEED 420
Qy 244 GYGFANDRIIAEDSGFVMSHGDHNYHFFKKDLTEEQIKAAQKHLVEVKTSHNGLDSLS 303
Db 421 GYGFANDRIIAEDSGFVMSHGDHNYHFFKKDLTEEQIKAAQKHLVEVKTSHNGLDSLS 480
Qy 304 HQDYPSNAKEMKDLKIEKIAIGIMKQYGVKRESIVVNNKRNIAIYPHGDHHDHADPID 363
Db 481 HQDYPSNAKEMKDLKIEKIAIGIMKQYGVKRESIVVNNKRNIAIYPHGDHHDHADPID 540
Qy 364 EHKPVGIGHSHSNVELKPEEGVAKKGNKYVTGEBELTNVNLKSTFNQNTFLANGQ 423
Db 541 EHKPVGIGHSHSNVELKPEEGVAKKGNKYVTGEBELTNVNLKSTFNQNTFLANGQ 600
Qy 424 KRVSFPPPELEKKLGINMLVKLITPDGKYLEKYSKGVFGEVGNIANFELDQYLP 483
Db 601 KRVSFPPPELEKKLGINMLVKLITPDGKYLEKYSKGVFGEVGNIANFELDQYLP 660
```


AAU84021

ID AAU84021 standard; Peptide; 1019 AA.

XX AC AAU84021;

XX DT 08-MAY-2002 (first entry)

XX DE Truncated variant of S. pneumoniae BVH-3, BVH-3M.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

KW pneumonia; streptococcal bacterial infection; mutant; mutin.

XX OS Streptococcus pneumoniae.

XX OS Synthetic.

XX PN WO200198334-A2.

XX PD 27-DEC-2001.

XX PF 19-JUN-2001; 2001WO-CA00908.

XX PR 20-JUN-2000; 2000US-212683P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX DR WPI; 2002-122272/16.

XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

XX PT epitope-bearing polypeptides, useful as vaccine components for treating

XX PT or preventing streptococcal infections such as otitis media,

XX PT meningitis, and bacteraemia

XX PS Example 1; Page -; 113pp; English.

XX CC The invention describes an isolated polypeptide (I) with 70-90%

XX CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

XX CC comprising (I) is useful for therapeutic or prophylactic treatment of

XX CC meningitis, otitis media, bacteraemia or pneumonia infection in an

XX CC individual susceptible to these disorders. (III) is also useful for

XX CC therapeutic or prophylactic treatment of any streptococcal bacterial

XX CC infection (e.g., caused by Streptococcus pneumoniae, group A

XX CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

XX CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocardia or

XX CC Staphylococcus aureus) in an individual susceptible to the infection.

XX CC A polynucleotide (III) encoding (I) is useful in DNA immunisation

XX CC techniques. The Streptococcus polypeptides are useful in a diagnostic

XX CC test for S. pneumoniae infection. (III) is useful for designing DNA

XX CC probes for use in detecting the presence of Streptococcus in a biological

XX CC sample suspected of containing the bacteria. The DNA probes may also be

XX CC used for detecting circulating S. pneumonia nucleic acid in a sample for

XX CC diagnosing streptococcal infections. This sequence represents a truncate

XX CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,

XX CC described in the method of the invention.

XX CC Note: This sequence does not appear in the specification but has

XX CC been created according to information given in the invention.

XX CC

XX SQ Sequence 1019 AA;

Query Match

Best Local Similarity 97.3%; Score 4276.5; DB 23; Length 1019;

Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNQRSEKDNRRSYVDGSSQSKSENTPQVSKQEGIAEQIVIKITDGGYV 60

DB 1 CAYALNQRSEKDNRRSYVDGSSQSKSENTPQVSKQEGIAEQIVIKITDGGYV 60

QY 61 TSHGDHYHYNGKPYDPALEFSEELMKDPNVLKADIVNEVGKGYIIVKVDGYIYVVKLD 120

DB 61 TSHGDHYHYNGKPYDPALEFSEELMKDPNVLKADIVNEVGKGYIIVKVDGYIYVVKLD 120

QY 121 AAHADNVRTKDEINROKQEHVKDNEKYNVAVARSQGRVTTINDGYVFNADIIEDTGA 180
 DB 121 AAHADNVRTKDEINROKQEHVKDNEKYNVAVARSQGRVTTINDGYVFNADIIEDTGA 180
 QY 181 YIVPHGGHYHYIPKSDLSASELAACAAHLAGKNMQPSQLSYSS----- 223
 DB 181 YIVPHGGHYHYIPKSDLSASELAACAAHLAGKNMQPSQLSYSS----- 223
 QY 224 ----- 223
 DB 241 PANKSENLOSLLKELYDSPAQRYSESGLVFPDPAKIIISRTPNGVAIPRHDHYHFPYSK 300
 QY 224 ----- 223
 DB 301 LSALEEKIARMVPISGTGVSTVSTNAKENEVVSSGLSSNPPSLTTSKELSSASDGYIEN 360
 QY 224 -----TPSPSLPINTPGTSHKHEED 243
 DB 361 PKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINTPGTSHKHEED 420
 QY 244 GYGFANDRIIAEDESQFVMSHGDHNYFFPKKDLTEQIKAAQKHLFEVKTSHNGLDSLSS 303
 DB 421 GYGFANDRIIAEDESQFVMSHGDHNYFFPKKDLTEQIKAAQKHLFEVKTSHNGLDSLSS 480
 QY 304 HEODYPSNAKEMKDLKKIEEKIAGIMKQYGVKRESIVVKNKNAIIYPHGDHHAADPTD 363
 DB 481 HEODYPSNAKEMKDLKKIEEKIAGIMKQYGVKRESIVVKNKNAIIYPHGDHHAADPTD 540
 QY 364 EHPVGIHSHSNYELFKPEGVAKKEGKNVYTGRIITVNVNLLKNSTNNQNTLANGQ 423
 DB 541 EHPVGIHSHSNYELFKPEGVAKKEGKNVYTGRIITVNVNLLKNSTNNQNTLANGQ 600
 QY 424 KRVSFSPPELEKKLGINMLVKLITPDGKVLKVSQKVGEGVGNIANFELDQPYLPQGT 483
 DB 601 KRVSFSPPELEKKLGINMLVKLITPDGKVLKVSQKVGEGVGNIANFELDQPYLPQGT 660
 QY 484 FKVTIASKDYPEVSYDGTFTVPTSLAYKASQTIFFYFFHAGDTYLRVNPQFAVPKGTDAL 543
 DB 661 FKVTIASKDYPEVSYDGTFTVPTSLAYKASQTIFFYFFHAGDTYLRVNPQFAVPKGTDAL 720
 QY 544 VRVDFEFGHNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTETMANAYLDNQSYIVE 603
 DB 721 VRVDFEFGHNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTETMANAYLDNQSYIVE 780
 QY 604 VPILEENQTDKPSILPQPKRKAQENSKLDERKVEPEPKTSEKVEKEKLSGTNSTNSTL 663
 DB 781 VPILEENQTDKPSILPQPKRKAQENSKLDERKVEPEPKTSEKVEKEKLSGTNSTNSTL 840
 QY 664 EEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSSGEVKKNNMADFTGEAPQGN 723
 DB 841 EEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSSGEVKKNNMADFTGEAPQGN 900
 QY 724 GENKPSNGKVSTGTVENOPTENKPADSLPEAPNEKVPKVPENSTDNGLNPNENGVSDPM 783
 DB 901 GENKPSNGKVSTGTVENOPTENKPADSLPEAPNEKVPKVPENSTDNGLNPNENGVSDPM 960
 QY 784 LDSALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVKKNL 837
 DB 961 LDSALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVKKNL 1014

RESULT 6

AAB12715

ID AAB12715 standard; Protein; 1039 AA.

XX AC AAB12715;

XX DT 21-NOV-2000 (first entry)

XX DE Streptococcus pneumoniae BVH-3 protein antigen SEQ ID NO:2.

XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

WO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

N-PSDB; AAA65730.

Streptococcal antigens useful for vaccinating against e.g. meningitis, otitis media, bacteremia and/or pneumonia

Claim 18; Fig 2; 106pp; English.

The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteremia and/or pneumonia. The present sequence represents the S. pneumoniae BVH-3 protein antigen.

Sequence 1039 AA;

Query Match 97.3%; Score 4276.5; DB 21; Length 1039;

Best Local Similarity 82.3%; Pred. No. 3.1e-254;

Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNQHRSQENKNNRYSYVDGSSQSKSENLPDQVSOKEGIAEQIVIKITDQGYV 60
 DB 21 CAYALNQHRSQENKNNRYSYVDGSSQSKSENLPDQVSOKEGIAEQIVIKITDQGYV 80
 QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNOLKADIVNEVKGYYIIKVDGKYVYLKD 120
 DB 81 TSHGDHYHYNGKVPYDALFSEELMKDPNOLKADIVNEVKGYYIIKVDGKYVYLKD 140
 QY 121 AAHADNVRTKDEINRQKQEHVKNKNSNVAARSQGRYTTNDGYVFPADIIEDTGN 180
 DB 141 AAHADNVRTKDEINRQKQEHVKNKNSNVAARSQGRYTTNDGYVFPADIIEDTGN 200
 QY 181 YIVPHGGHYHYIPKSDLSASELAHAKAHLAGNMOPSOISLSS 223
 DB 201 YIVPHGGHYHYIPKSDLSASELAHAKAHLAGNMOPSOISLSS 260
 QY 224 223
 DB 261 PANKSENLOSLLKELYDPSAQRYSESGLVDPKAKIISRTPNGVAIPHGDHYHPIYSK 320
 QY 224 223
 DB 321 LSALAEKIAARMVPIGCTGVSTNAKPNNEWSSGLSSNPSLTSKELSSASDGYIFN 380
 QY 224 223
 DB 381 PKDIVETATATVIRGHDHFHYIPKSNQIQPTLNNSLATPSPLPNPGTSHKHEED 440
 QY 244 GYGFDAHRIIAEDESQFVMSHGDHNNHYFFKDLTEEQKAAQKHLEEVKTSNGLDLSS 303
 DB 441 GYGFDAHRIIAEDESQFVMSHGDHNNHYFFKDLTEEQKAAQKHLEEVKTSNGLDLSS 500
 QY 304 HEQDYPSNAKEMKOLDKKEEIKAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPID 363

Db 501 HEQDYPSNAKEMKOLDKKEEIKAGIMKQYGVKRESIVVKNKNAIYPHGDHHDADPID 560
 QY 364 EHKPVGIGHSHSNYELFKPEEGVAKKEGKNKVTGEBLTNNVNLKKNSTNNQNETLANQ 423
 Db 561 EHKPVGIGHSHSNYELFKPEEGVAKKEGKNKVTGEBLTNNVNLKKNSTNNQNETLANQ 620
 QY 424 KRVSFPPPELEKRLKGINMLVLIITPDGKVLKVSQKVEGEGVGNIANFELDQPYLPQGT 483
 Db 621 KRVSFPPPELEKRLKGINMLVLIITPDGKVLKVSQKVEGEGVGNIANFELDQPYLPQGT 680
 QY 484 FKTIASKOYPEVSYDGTFTVPTSLAYKASQITFYFFHAGDTYLRVNPQFAVPKGTDL 543
 Db 681 FKTIASKOYPEVSYDGTFTVPTSLAYKASQITFYFFHAGDTYLRVNPQFAVPKGTDL 740
 QY 544 VRVDFEHGHNAYLNNYKVGEEKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSYIVE 603
 Db 741 VRVDFEHGHNAYLNNYKVGEEKLPIPKLNOGTRTAGNKIPVTFMANAYLDNQSYIVE 800
 QY 604 VPILKENQTDKPSILPQFARNKAQENSKLDERVEPKTSEKVEKEKLEKSETGNSTNSL 563
 Db 801 VPILKENQTDKPSILPQFARNKAQENSKLDERVEPKTSEKVEKEKLEKSETGNSTNSL 860
 QY 564 EEPVTVDPQVEKVAFAESYGMKLENNVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 723
 Db 861 EEPVTVDPQVEKVAFAESYGMKLENNVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
 QY 724 GENKPSGKSVSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVSGDPM 783
 Db 921 GENKPSGKSVSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPEGNVSGDPM 980
 QY 784 LDSALEAPADVPQVEKLEKFTASYGLGDSVIFNMDGTIELRLPSEVTKKL 837
 Db 981 LDSALEAPADVPQVEKLEKFTASYGLGDSVIFNMDGTIELRLPSEVTKKL 1034

RESULT 7

AAU75932

ID AAU75932 standard; Protein; 1039 AA.

AC AAU75932;

DT 08-MAY-2002 (first entry)

DE Streptococcus pneumoniae BVH-3 protein version #1.

XX BVH-3; BVH-11; vaccine; meningitis; Otitis media; bacteraemia;

XX pneumonia; streptococcal bacterial infection.

OS Streptococcus pneumoniae.

PN WO200198334-A2.

PD 27-DEC-2001.

PF 19-JUN-2001; 2001WO-CA00908.

PR 20-JUN-2000; 2000US-212683P.

PA (SHIR-) SHIRE BIOCHEM INC.

PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX WPI; 2002-122272/15.

DR N-PSDB; ABK15101.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and

PT epitope-bearing polypeptides, useful as vaccine components for treating

PT or preventing streptococcal infections such as otitis media,

XX meningitis, and bacteraemia

PS Example 1; Fig 6; 113pp; English.

XX

CC The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This is the amino acid sequence of
CC Streptococcus pneumoniae protein BVH-3, used to create the antigenic
CC peptides described in the method of the invention.

xx Sequence 1039 AA;

Query Match 97.3%; Score 4276.5; DB 23; Length 1039;
Best Local Similarity 82.3%; Pred. No. 3.1e-254;

Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNQRHSEKNNRVSVYDGSQSKSENTPDOVSQKEGQIAEQIVIKITDQGV 60
DB 21 CAYALNQRHSEKNNRVSVYDGSQSKSENTPDOVSQKEGQIAEQIVIKITDQGV 80
QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGGYIIVKDGKYYVYLKD 120
DB 81 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGGYIIVKDGKYYVYLKD 140
QY 121 AAHADNVRTKDEINQKQEHVKDNEKVNNAVARSQRYTNDGYVFNPAIDIEDTGN 180
DB 141 AAHADNVRTKDEINQKQEHVKDNEKVNNAVARSQRYTNDGYVFNPAIDIEDTGN 200
QY 181 YIVPHGHHYIPKSDLSASLAAKAHLAGKNQPSQLSYSS----- 223
DB 201 YIVPHGHHYIPKSDLSASLAAKAHLAGKNQPSQLSYSS----- 260
QY 224 ----- 223
DB 261 PANKSENLOLLKELYDSPAQRVSESDGLVFPDPAKIISRTPNGVALPHGDHYHIFYSK 320
QY 224 ----- 223
DB 321 LSALAEKIARWVPISGTGSTVSTNAKNPEVSVSLGSLSSNPSSLTTSKELSSASDGVIFN 380
QY 224 -----TPSPSLPINPCTSHKHEED 243
DB 381 PKDIVVEETATAYIVRHGDHFFHYIPKSNQIGOPTLPNNSLATPSPSLPINPCTSHKHEED 440
QY 244 GYGFDANRIIAEDESFGVMSGHDHNYEFKKDLTEEQKAAQKHLKEEVKTSNHLGDSLS 303
DB 441 GYGFDANRIIAEDESFGVMSGHDHNYEFKKDLTEEQKAAQKHLKEEVKTSNHLGDSLS 500
QY 304 HQDDYPSNAKEMKLDLKEEKIAGIMKQYGVKRESIYVWNEKNAIYPHGDHHAAPID 363
DB 501 HQDDYPSNAKEMKLDLKEEKIAGIMKQYGVKRESIYVWNEKNAIYPHGDHHAAPID 560
QY 364 EHKPVGIGHSNVELFPPEGVAKKNGKYYTGEELTNVYVNLKNSFTNQNFTLANGQ 423
DB 561 EHKPVGIGHSNVELFPPEGVAKKNGKYYTGEELTNVYVNLKNSFTNQNFTLANGQ 620
QY 424 KRVSFPFPELEKLGINMLVKLITPDGKVLKYSKGVFGVGVGNIANFELDQPYLPQGT 483
DB 621 KRVSFPFPELEKLGINMLVKLITPDGKVLKYSKGVFGVGVGNIANFELDQPYLPQGT 680
QY 484 FKYTITASKDYPEVSDGTFTTPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 543

DB 681 FKYTITASKDYPEVSDGTFTTPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 740
QY 544 VRVEDEHGNAYLENNYKVGEIKLPIPKLNOGTRTAGNKIPVTEMANAYLDNOSTYIVE 603
DB 741 VRVEDEHGNAYLENNYKVGEIKLPIPKLNOGTRTAGNKIPVTEMANAYLDNOSTYIVE 800
QY 504 VPILEKENQTDKPSILPQFRKRAQENSKLDEKVEEPEKTSKVEKEKLSGTGNSNSTL 663
DB 801 VPILEKENQTDKPSILPQFRKRAQENSKLDEKVEEPEKTSKVEKEKLSGTGNSNSTL 860
QY 664 EEVPTVPDQVEKVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNMADFTGEAPQGN 723
DB 861 EEVPTVPDQVEKVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNMADFTGEAPQGN 920
QY 724 GENKPSNGKVGSTGVNQPTENKPADSLPEAPNEKPVKPNENSTDGMLNPEGNVGSDDPM 783
DB 921 GENKPSNGKVGSTGVNQPTENKPADSLPEAPNEKPVKPNENSTDGMLNPEGNVGSDDPM 980
QY 784 LDSALEAPAVDPVOEKLKFTASYGLGDSVIFNMDGTIELRLPDSGEVIRKKNL 837
DB 981 LDPALEAPAVDPVOEKLKFTASYGLGDSVIFNMDGTIELRLPDSGEVIRKKNL 1034

RESULT 8
AAB12750

ID AAB12750 standard; Protein; 1019 AA.

XX AC AAB12750;

XX DT 21-NOV-2000 (first entry)

XX DE Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.

XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX KW Prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

XX OS Streptococcus pneumoniae; immunisation; bactericidal.

XX OS Streptococcus pneumoniae.

XX PN W0200039299-A2.

XX PD 06-JUL-2000.

XX PF 20-DEC-1999; 99WO-CA01218.

XX PR 23-DEC-1998; 98US-0113800.

XX PA (BIOC-) BIOCHEM PHARMA INC.

XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

XX PS WPI; 2000-452397/39.

XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,

XX PT otitis media, bacteraemia and/or pneumonia

XX PS Disclosure; Fig 11; 106pp; English.

CC The present invention describes nucleic acids (I) encoding protein

CC antigens (II) from Streptococcus pneumoniae. The protein antigens

CC have bactericidal activity. The nucleic acids, encoding the protein

CC antigens, may be used for the recombinant production of the proteins

CC they encode. The protein antigens may then be used as vaccines for the

CC prevention and treatment of Streptococcal infections in mammals

CC (especially humans) which result in, e.g. meningitis, otitis media,

CC bacteraemia and/or pneumonia. The present sequence represents a

CC S. pneumoniae BVH-3 protein antigen, from the present invention.

XX SQ Sequence 1019 AA;

Query Match 97.2%; Score 4274.5; DB 21; Length 1019;

Best Local Similarity 82.3%; Pred. No. 4.1e-254;

Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;


```

Db 361 PKDIVEETATAYIVRGDHFHYTPKSNQIQPTLPNNSLATPSPSLINPGTSHKHEED 420
QY 244 GYGFDANRIIAEDSGFVMSHGDHNYFFFKKDLTEQIKAAQKHLEEVKTSNGLDLSL 303
Db 421 GYGFDANRIIAEDSGFVMSHGDHNYFFFKDLTEQIKAAQKHLEEVKTSNGLDLSL 480
QY 304 HEQDYPNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVNKEKNALIIYPHGDHHAADPD 363
Db 481 HEQDYPNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVNKEKNALIIYPHGDHHAADPD 540
QY 364 EHPVIGHSNHYELFKPEEGVAKKGNKVTGELTNVNVLLKKNSTFNNQNTLANGQ 423
Db 541 EHPVIGHSNHYELFKPEEGVAKKGNKVTGELTNVNVLLKKNSTFNNQNTLANGQ 600
QY 424 KRVSFPPEPELEKLGINMLVLIITPDGVLEKVSQKVFGEVGNIANFELDQYLPQGT 483
Db 601 KRVSFPPEPELEKLGINMLVLIITPDGVLEKVSQKVFGEVGNIANFELDQYLPQGT 660
QY 484 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPGKTDAL 543
Db 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPGKTDAL 720
QY 544 VRVDFEFGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 603
Db 721 VRVDFEFGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 780
QY 604 VPILKENOTDKPSILPQPKRKAQNSKLDKVEEPKTSKVEKEKLSGTGNTSNTL 663
Db 781 VPILKENOTDKPSILPQPKRKAQNSKLDKVEEPKTSKVEKEKLSGTGNTSNTL 840
QY 664 EYVPTVDPQVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQGN 723
Db 841 EYVPTVDPQVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQGN 900
QY 724 GENKPSGKSVGTVENOPTENKPADSLPEAPNEXPKPENSTDNGLNPEGVGSDPM 783
Db 901 GENKPSGKSVGTVENOPTENKPADSLPEAPNEXPKPENSTDNGLNPEGVGSDPM 960
QY 784 LDSALEEAPADVPQVKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVKKNL 837
Db 961 LDPALAEAPADVPQVKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVKKNL 1014

```

RESULT 10

```

AAB12752
ID AAB12752 standard; Protein; 1019 AA.
AC
XX
AC AAB12752;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain P4241 BVH-3 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
DR

```

```

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
XX Disclosure; Fig 11; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

```

Query Match 97.1%; Score 4268.5; DB 21; Length 1019;

Best Local Similarity 82.2%; Pred. No. 9.5e-254;

Matches 834; Conservative 0; Mismatches 3; Indels 177; Gaps 1;

QY 1 CAYALNQHSQENKNNRVSYVDGSSQSKSENLTPOVSOKEGIAEQIVIKITDQYV 60

Db 1 CAYALNQHSQENKNNRVSYVDGSSQSKSENLTPOVSOKEGIAEQIVIKITDQYV 60

QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLDADIVNEVKGYIITKVDGKYVYVLUK 120

Db 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLDADIVNEVKGYIITKVDGKYVYVLUK 120

QY 121 AAHADNVRTKDEINRQKQEHVKDNEKVNNAVAVASQGRYTTNDGYVFNPAIIEEDTGN 180

Db 121 AAHADNVRTKDEINRQKQEHVKDNEKVNNAVAVASQGRYTTNDGYVFNPAIIEEDTGN 180

QY 181 YIVPHGGHYHYLPKSDLSASELAHAKHACKNMOPSOISYSS 223

Db 181 YIVPHGGHYHYLPKSDLSASELAHAKHACKNMOPSOISYSS 223

QY 241 PANKSENQLSKLLELYDPSAQRYSESDGLVDPAPAKIISRTPNGVAIPHGDIHFIPYSK 300

Db 241 PANKSENQLSKLLELYDPSAQRYSESDGLVDPAPAKIISRTPNGVAIPHGDIHFIPYSK 300

QY 224 223

Db 224 223

QY 301 LSALLEEKIARMVPIISGTSTVSTNAKPNEVSSLSGSSNPFSLTTSKELSSASDGYIFN 360

Db 301 LSALLEEKIARMVPIISGTSTVSTNAKPNEVSSLSGSSNPFSLTTSKELSSASDGYIFN 360

QY 224 243

Db 224 243

QY 361 PKDIVEETATAYIVRGDHFHYTPKSNQIQPTLPNNSLATPSPSLINPGTSHKHEED 420

Db 361 PKDIVEETATAYIVRGDHFHYTPKSNQIQPTLPNNSLATPSPSLINPGTSHKHEED 420

QY 244 GYGFDANRIIAEDSGFVMSHGDHNYFFFKKDLTEQIKAAQKHLEEVKTSNGLDLSL 303

Db 244 GYGFDANRIIAEDSGFVMSHGDHNYFFFKKDLTEQIKAAQKHLEEVKTSNGLDLSL 303

QY 304 HEQDYPNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVNKEKNALIIYPHGDHHAADPD 363

Db 304 HEQDYPNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVNKEKNALIIYPHGDHHAADPD 363

QY 481 HEQDYPNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVNKEKNALIIYPHGDHHAADPD 540

Db 481 HEQDYPNAKEMKDLKKTEETAGTAKMKGQYGVKRESIVNKEKNALIIYPHGDHHAADPD 540

QY 364 EHPVIGHSNHYELFKPEEGVAKKGNKVTGELTNVNVLLKKNSTFNNQNTLANGQ 423

Db 364 EHPVIGHSNHYELFKPEEGVAKKGNKVTGELTNVNVLLKKNSTFNNQNTLANGQ 423

QY 541 EHPVIGHSNHYELFKPEEGVAKKGNKVTGELTNVNVLLKKNSTFNNQNTLANGQ 600

Db 541 EHPVIGHSNHYELFKPEEGVAKKGNKVTGELTNVNVLLKKNSTFNNQNTLANGQ 600

QY 424 KRVSFPPEPELEKLGINMLVLIITPDGVLEKVSQKVFGEVGNIANFELDQYLPQGT 483

Db 424 KRVSFPPEPELEKLGINMLVLIITPDGVLEKVSQKVFGEVGNIANFELDQYLPQGT 483

QY 601 KRVSFPPEPELEKLGINMLVLIITPDGVLEKVSQKVFGEVGNIANFELDQYLPQGT 660

Db 601 KRVSFPPEPELEKLGINMLVLIITPDGVLEKVSQKVFGEVGNIANFELDQYLPQGT 660

QY 484 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPGKTDAL 543

Db 484 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPGKTDAL 543

QY 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPGKTDAL 720

Db 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPPFHAGDTYLRVNPQFAVPGKTDAL 720

QY 544 VRVDFEFGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 603

Db 544 VRVDFEFGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 603

QY 721 VRVDFEFGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 780

Db 721 VRVDFEFGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 780

```
QY 604 VPILEKENOTDKPSILPOFKRNKAQENSKLDEKVEPKTSEKVEKEKILSETGNSNSTL 663
DB 781 VPILEKENOTDKPSILPOFKRNKAQENSKLDEKVEPKTSEKVEKEKILSETGNSNSTL 840
QY 664 EEVPTVDPVQEKVAFESYGMKLENVLFNMDGTIELVLPSCGEVTKKKNMADFTGEAPOGN 723
DB 841 EEVPTVDPVQEKVAFESYGMKLENVLFNMDGTIELVLPSCGEVTKKKNMADFTGEAPOGN 900
QY 724 GENKPSENGKVSSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTONGMLNPEGNGVSDPM 783
DB 901 GENKPSENGKVSSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTONGMLNPEGNGVSDPM 960
QY 784 LSALBEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGEVVKKNL 837
DB 961 LDPALBEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGEVVKKNL 1014

RESULT 11
AAB12753
ID AAB12753 standard; Protein; 1019 AA.
XX AC AAB12753;
XX 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae strain A66 BVH-3 protein antigen.
XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX KW otitis media; pneumonia; immunisation; bactericidal.
XX OS Streptococcus pneumoniae.
XX PN WO200039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA01218.
XX PR 23-DEC-1998; 98US-0113800.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteraemia and/or pneumonia.
XX PS Disclosure; Fig 11; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens
XX CC have bactericidal activity. The nucleic acids, encoding the protein
XX CC antigens, may be used for the recombinant production of the proteins
XX CC they encode. The protein antigens may then be used as vaccines for the
XX CC prevention and treatment of Streptococcal infections in mammals
XX CC (especially humans) which result in, e.g. meningitis, otitis media,
XX CC bacteraemia and/or pneumonia. The present sequence represents a
XX CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX 50 Sequence 1019 AA;

Query Match 97.1%; Score 4268.5; DB 21; Length 1019;
Best Local Similarity 82.2%; Pred. No. 9.5e-254;
Matches 834; Conservative 0; Mismatches 23; Indels 177; Gaps 1;

QY 1 CAYALNQRHSQENKDNRRVSYVDGSSQSKSENLTPOVSKQEGIQAEQIVKITDQGYV 60
DB 1 CAYALNQRHSQENKDNRRVSYVDGSSQSKSENLTPOVSKQEGIQAEQIVKITDQGYV 60
```

```
QY 61 TSHGDHYHYNGKVPYDALFSEELLKMKDPYQLKDADIVNEVKGYYIIKVDGKYYVYLKD 120
DB 61 TSHGDHYHYNGKVPYDALFSEELLKMKDPYQLKDADIVNEVKGYYIIKVDGKYYVYLKD 120
QY 121 AAHADNVRTKDEINRQKOEHVKNKNSVNAVARSOGRYTTNDGYVFNPAADIIEDTGN 180
DB 121 AAHADNVRTKDEINRQKOEHVKNKNSVNAVARSOGRYTTNDGYVFNPAADIIEDTGN 180
QY 181 YIVPHGCHYHYIPKSDLSASELAALAAKAGKKNQPSOLSYSS----- 223
DB 181 YIVPHGCHYHYIPKSDLSASELAALAAKAGKKNQPSOLSYSS----- 223
QY 241 PANKSENLSLLKELYDPSAQRYSSEGLVFPDAKIIISRTPNGVAIPHGDHYEFIPYSK 300
DB 241 PANKSENLSLLKELYDPSAQRYSSEGLVFPDAKIIISRTPNGVAIPHGDHYEFIPYSK 300
QY 301 LSALBEKIARWPIISGTGSTVSTNAKNPNEVVSSLSGSSNPSSLTTSKELSSASDGYIFN 360
DB 301 LSALBEKIARWPIISGTGSTVSTNAKNPNEVVSSLSGSSNPSSLTTSKELSSASDGYIFN 360
QY 361 PKDIVEETATAYIVRHGDHPHYIPKSNQIGOPTLPNNSLATPSPSLPINPSTSEKHEED 420
DB 361 PKDIVEETATAYIVRHGDHPHYIPKSNQIGOPTLPNNSLATPSPSLPINPSTSEKHEED 420
QY 421 GYGFANRIIAEDESFGFVMSHGDHNYFFKKDLTEEOIKAAQKHLEEVKTSNGLDLSLS 303
DB 421 GYGFANRIIAEDESFGFVMSHGDHNYFFKKDLTEEOIKAAQKHLEEVKTSNGLDLSLS 303
QY 304 HEQYDPSNAKEMKOLDKKIEBKTAGINKQYGVKRESIVNKEKNALIIYPHGDHHDAPID 363
DB 304 HEQYDPSNAKEMKOLDKKIEBKTAGINKQYGVKRESIVNKEKNALIIYPHGDHHDAPID 363
QY 481 HEQYDPSNAKEMKOLDKKIEBKTAGINKQYGVKRESIVNKEKNALIIYPHGDHHDAPID 540
DB 481 HEQYDPSNAKEMKOLDKKIEBKTAGINKQYGVKRESIVNKEKNALIIYPHGDHHDAPID 540
QY 364 EHKPVGICHSHSNYELFKPDEGVAKKGNKYVTGEBELTNVNNLLKNTFNQNFILANG 423
DB 364 EHKPVGICHSHSNYELFKPDEGVAKKGNKYVTGEBELTNVNNLLKNTFNQNFILANG 423
QY 541 EHKPVGICHSHSNYELFKPDEGVAKKGNKYVTGEBELTNVNNLLKNTFNQNFILANG 600
DB 541 EHKPVGICHSHSNYELFKPDEGVAKKGNKYVTGEBELTNVNNLLKNTFNQNFILANG 600
QY 424 KRVSFSPPELEKLGINLVKLTTPDGKVLKESGKVGEGVGNIANFELDQYLPFGQT 483
DB 424 KRVSFSPPELEKLGINLVKLTTPDGKVLKESGKVGEGVGNIANFELDQYLPFGQT 483
QY 601 KRVSFSPPELEKLGINLVKLTTPDGKVLKESGKVGEGVGNIANFELDQYLPFGQT 660
DB 601 KRVSFSPPELEKLGINLVKLTTPDGKVLKESGKVGEGVGNIANFELDQYLPFGQT 660
QY 484 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAVPGKTDAL 543
DB 484 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYPFHAGDTYLRVNPQFAVPGKTDAL 543
QY 544 VEVDFEFHGNAYLENNYKVGELKPIPKLNOGTTTAGNKIPVTFMANAYLDNGSTIVE 603
DB 544 VEVDFEFHGNAYLENNYKVGELKPIPKLNOGTTTAGNKIPVTFMANAYLDNGSTIVE 603
QY 721 VRVDFEFHGNAYLENNYKVGELKPIPKLNOGTTTAGNKIPVTFMANAYLDNGSTIVE 780
DB 721 VRVDFEFHGNAYLENNYKVGELKPIPKLNOGTTTAGNKIPVTFMANAYLDNGSTIVE 780
QY 604 VPILKENEQTDKPSILPOFKRNKAQENSKLDEKVEPKTSEKVEKEKILSETGNSNSTL 663
DB 604 VPILKENEQTDKPSILPOFKRNKAQENSKLDEKVEPKTSEKVEKEKILSETGNSNSTL 663
QY 664 EEVPTVDPVQEKVAFESYGMKLENVLFNMDGTIELVLPSCGEVTKKKNMADFTGEAPOGN 723
DB 664 EEVPTVDPVQEKVAFESYGMKLENVLFNMDGTIELVLPSCGEVTKKKNMADFTGEAPOGN 723
QY 841 EEVPTVDPVQEKVAFESYGMKLENVLFNMDGTIELVLPSCGEVTKKKNMADFTGEAPOGN 900
DB 841 EEVPTVDPVQEKVAFESYGMKLENVLFNMDGTIELVLPSCGEVTKKKNMADFTGEAPOGN 900
QY 724 GENKPSENGKVSSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTONGMLNPEGNGVSDPM 783
DB 724 GENKPSENGKVSSTGTVENOPTENKPADSLPEAPNEKPKVPKPNSTONGMLNPEGNGVSDPM 783
QY 784 LSALBEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGEVVKKNL 837
DB 784 LSALBEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGEVVKKNL 837
QY 961 LDPALBEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGEVVKKNL 1014
DB 961 LDPALBEAPAVDPVQEKLEKFTASVGLGLDSVIFNMDGTIELRLPSGEVVKKNL 1014

RESULT 12
AAB12749
ID AAB12749 standard; Protein; 1019 AA.
XX AC AAB12749;
XX 21-NOV-2000 (first entry)
XX DT 21-NOV-2000 (first entry)
XX XX
```


DE Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
OS Streptococcus pneumoniae.
XX
XX WO200039299-A2.
XX
PD 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
PF
XX 23-DEC-1998; 98US-0113800.
PR
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
XX WPI; 2000-452397/39.
DR
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia
XX
PS Disclosure; Fig 11; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins.
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

Query Match 97.0%; Score 4264.5; DB 21; Length 1019;
Best Local Similarity 82.1%; Pred. No. 1.7e-253;
Matches 832; Conservative 2; Mismatches 3; Indels 177; Gaps 1;

Qy 1 CAYALNQHRSENKNNRVSYVDGSSQSKSENLTPOVSKQEGTQABQIVIKITDGYV 60
Db 1 CAYALNQHRSENKNNRVSYVDGSSQSKSENLTPOVSKQEGTQABQIVIKITDGYV 60
Qy 61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQLKDADIVNEVKGGYIKYDGYVYLKD 120
Db 61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQLKDADIVNEVKGGYIKYDGYVYLKD 120
Qy 121 AAHADNVRTKDEINRQKQEHVKDKENKNSVAVARSQGRYTTNDGYVFNPAIDIEDTGA 180
Db 121 AAHADNVRTKDEINRQKQEHVKDKENKNSVAVARSQGRYTTNDGYVFNPAIDIEDTGA 180
Qy 181 YIVPHGGHYHYIPKSDLSASELAALAHLAGKNMOPQSOLSYSS----- 223
Db 181 YIVPHGGHYHYIPKSDLSASELAALAHLAGKNMOPQSOLSYSSSTASDNNTOSVAKGSTK 240
Qy 224 ----- 223
Db 241 PANKSENLOSLLKELYDSPSAQRYSSEGLVDPDAKIISRTPNPVAIPHGDHYHFIPYSK 300
Qy 224 ----- 223
Db 301 LSALBEKIARRVPISTGTSTYTNKAPNEVYSSLSLSSNPFSSLTTSKELSSASDGYFN 360
Qy 224 -----TPSPSLPINPOTSHEKHEED 243
Db 361 PADIIVEETATYIVVHGDHYHYIPKSNQIGOPTLPNNSLATPSPSLPINPOTSHEKHEED 420
Qy 244 GYGFANRIIAEDSGFIMSHGNHNYFFKKDLTTEQIKAAQKHLSEVKTSHNGLDSLSS 303
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

Db 421 GYGFANRIIAEDSGFIMSHGNHNYFFKKDLTTEQIKAAQKHLSEVKTSHNGLDSLSS 480
Qy 304 HEQDYPNNAEMKMDLKKIEEKTAGINKQYGVKRESIVVKNKNALIIYPGDDHHHADPID 363
Db 481 HEQDYPNNAEMKMDLKKIEEKTAGINKQYGVKRESIVVKNKNALIIYPGDDHHHADPID 540
Qy 364 EHKPVGIGCHSHSNYELFKPEEGYAKKEGNKVYTGEEELTNVYVNLKKNSTENNQNFLLANGQ 423
Db 541 EHKPVGIGCHSHSNYELFKPEEGYAKKEGNKVYTGEEELTNVYVNLKKNSTENNQNFLLANGQ 600
Qy 424 KRVSFSPPELEKKGINMLVKLITPDGKVLKVKSVKGVGEGVGNANTANFELDQYLPQGQT 483
Db 601 KRVSFSPPELEKKGINMLVKLITPDGKVLKVKSVKGVGEGVGNANTANFELDQYLPQGQT 660
Qy 484 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAYPKGTDAL 543
Db 661 FKYTIAASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAYPKGTDAL 720
Qy 544 VRFDFPHGNAYLENNYKVGGEIKLPIPKLNOGTTRAGNKIPVTFMANAYLDNOSTYIVE 603
Db 721 VRFDFPHGNAYLENNYKVGGEIKLPIPKLNOGTTRAGNKIPVTFMANAYLDNOSTYIVE 780
Qy 604 VPILLEKENQDKPSILPQFRKNAQENSKLDEKVEBPKTSEKVEKEKLSGTGNSNSTL 663
Db 781 VPILLEKENQDKPSILPQFRKNAQENSKLDEKVEBPKTSEKVEKEKLSGTGNSNSTL 840
Qy 664 EEVPTVPDQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIRKKNMADFTGEAPQGN 723
Db 841 EEVPTVPDQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIRKKNMADFTGEAPQGN 900
Qy 724 GENKPSNGKSVGTGTVENQNTENKPADSLPEAPNEKPVKPESTDNQMLNPEGNVGSDDPM 783
Db 901 GENKPSNGKSVGTGTVENQNTENKPADSLPEAPNEKPVKPESTDNQMLNPEGNVGSDDPM 960
Qy 784 LDSALEAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSEGEVIRKKNL 837
Db 961 LDPALAPAVDPVQEKLEKFTASYGLGDSVIFNMDGTIELRLPSEGEVIRKKNL 1014

RESULT 13
AAU83997
ID AAU83997 standard; Peptide; 807 AA.
XX
AC AAU83997;
XX
DT 08-MAY-2002 (first entry)
XX
DE Truncated variant of S. pneumoniae BVH-11, NEW25.
XX
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.
XX
OS Streptococcus pneumoniae.
OS Synthetic.
XX
XX WO200198334-A2.
PN
XX 27-DEC-2001.
XX
PF 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
PR
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX WPI; 2002-122272/16.
DR
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

XX PS Example 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus or Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumoniae nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a Streptococcus pneumoniae gene used to obtain antigenic peptides, described in the method of the invention.

CC Note: This sequence does not appear in the specification but has been created according to information given in the invention.

XX SQ Sequence 807 AA;

Query Match 77.4%; Score 3402.5; DB 23; Length 807;
 Best Local Similarity 79.6%; Pred. No. 1.1e-200;
 Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11:

QY 10 SOENKNNRVYVYDGSQ--SOKSEN1-----TPDQVSKQEGTQAEQI-VIKIT 55
 DB 10 SSTASDNTQVAKGTSKPANKSENLSLLKELYDPSAQRYSSEGLVDFDPAKIISRT 69

QY 56 DOGYVTSHGHDHYHNGKVPYDALPS-BELLMKDPNYQLKDADIVNEKVGVIIVKVDGY 114
 DB 70 PNGVAIPGHDHFF-----IPYSKLSALEEKIAR-----MVPISGT- 105

QY 115 YVYLKDAAHADNRVTKDEINRQKHVKDNKVNMSVAVARSQGRYTT-----NDGYV 167
 DB 106 -----GSTVST-----NAKNEVVSLSGLSSNPSSILTSKELSSASDGYI 146

QY 168 FNPADIIETGNAYIVPHGCHVHYTPKSDLSASELAAKAHLAKGNMOPSQLSYS-STPS 226
 DB 147 FNPKDIVEETATAYIVRHGDHFHYTPKSNQIG-----OPTLFNNSLATPS 191

QY 227 PSLPIMPNGTSHEKHEDGSGFDANRTIAEDSGFVMSHGDHNYFFKDLTEEQIKAAQK 286
 DB 192 PSLPIMPNGTSHEKHEDGSGFDANRTIAEDSGFVMSHGDHNYFFKDLTEEQIKAAQK 251

QY 287 HLBVKTSHNGLDLSLSSHQDYPYNSAKEMKDLKDKIEEKIAGIMQYGVKRESIYVNNKEK 346
 DB 252 HLBVKTSHNGLDLSLSSHQDYPYNSAKEMKDLKDKIEEKIAGIMQYGVKRESIYVNNKEK 311

QY 347 NAIYPHGDHHDADPDEKHPYGVIGHSHSNYELFRPEEGVAKKEGKNVYTGELTNVNL 406
 DB 312 NAIYPHGDHHDADPDEKHPYGVIGHSHSNYELFRPEEGVAKKEGKNVYTGELTNVNL 371

QY 407 LKNSFTNNQNTLANGOKRVSFSPPELEKLGINMLVKLIITPDGKVLKESGKVFGEV 466
 DB 372 LKNSFTNNQNTLANGOKRVSFSPPELEKLGINMLVKLIITPDGKVLKESGKVFGEV 431

QY 467 GNTANPELOQPLPGGTFFYTTASKDYPYVSDGTFTVPTSLAYKMASOTIFYPPHAGDT 526
 DB 432 GNTANPELOQPLPGGTFFYTTASKDYPYVSDGTFTVPTSLAYKMASOTIFYPPHAGDT 491

QY 527 YLRVNPQFAVPKGTDALVRVDFEFGHGNAYLNNYKVGELKLPKLNQCTTTAGNKIPV 586
 DB 492 YLRVNPQFAVPKGTDALVRVDFEFGHGNAYLNNYKVGELKLPKLNQCTTTAGNKIPV 551

QY 587 TFMANAYLDNQSTYIVVEPILKEKNTQDKPSILPOFKRNKAQENSKLDEKVEEPTSEKV 646

DB 552 TFMANAYLDNQSTYIVVEPILKEKNTQDKPSILPOFKRNKAQENSKLDEKVEEPTSEKV 611

QY 647 EKEKLSGTGNTSNTLEEVPTVDPOEKVAKFAESVGMKLENVLFNMDGTIELYLPSPGE 706
 DB 612 EKEKLSGTGNTSNTLEEVPTVDPOEKVAKFAESVGMKLENVLFNMDGTIELYLPSPGE 671

QY 707 VIKKNMADFTGEAPQNGENKPSGKNGVSTGTVENOPTENKPADSLPEAPNEKPVKPKNS 766
 DB 672 VIKKNMADFTGEAPQNGENKPSGKNGVSTGTVENOPTENKPADSLPEAPNEKPVKPKNS 731

QY 767 TDNGMLNPEGVSDPMLDSALEEAPAVDPVQKLEKFTASYGLGLDSVIFNMDGTIELR 826
 DB 732 TDNGMLNPEGVSDPMLDSALEEAPAVDPVQKLEKFTASYGLGLDSVIFNMDGTIELR 791

QY 827 LPSGEVIRKKNL 837

DB 792 LPSGEVIRKKNL 802

RESULT 14
 AAU84054
 ID AAU84054 standard; Peptide: 1152 AA.
 XX AC AAU84054;
 XX DT 08-MAY-2002 (first entry)
 XX DE S. pneumoniae derived chimeric peptide, NEW27.
 XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; muten;
 KW BVH-11-2.
 XX OS Streptococcus pneumoniae.
 XX OS Synthetic.
 XX FN WO200198334-A2.
 XX PD 27-DEC-2001.
 XX 19-JUN-2001; 2001WO-CA00908.
 XX 20-JUN-2000; 2000US-212683P.
 XX (SHIR-) SHIRE BIOCHEM INC.
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 WPI: 2002-122272/16.
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 epitope-bearing polypeptides, useful as vaccine components for treating
 or preventing streptococcal infections such as otitis media,
 meningitis, and bacteraemia -
 Example 1; Page -: 113pp; English.
 The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus or Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological

CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
SQ

Sequence 1152 AA;

Query Match 77.4%; Score 3402.5; DB 23; Length 1152;
Best Local Similarity 79.6%; Pred. No. 1.7e-200;
Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11;

QY 10 SQENKDNRVSYVDSQS--SQKSENL-----TPDQVSKQEGIQAEQI-VIKIT 55
DB 355 SSTAASDNTQSVAKGTSKPAKNSLSLKLKELYDSPAQRYSDESLGDFDPAKIIISRT 414
QY 56 DQGYVTSRHDHYHYNGVPYDALPS-BELLMKDPNYOLKADIVNEVKGYYIIKVDGKY 114
DB 415 PNGVAIPGHGHHYF---IPYKLSALEKTIAR-----MVPISGT- 450
QY 115 YYLKDAHADNVRTKDEINRQKQEHVXDNEKYNVNAVARSQGRYIT-----NDGYV 167
DB 451 -----GSTVST-----NAKPEVYVSLGSLSNPSSLITSKELSSASDGYI 491
QY 168 FNPADIIEDTGNAYIVPHGGHYHYTPKSDLSASELAATAKAGLAGKNMOPSOISYS-STPS 226
DB 492 FNPKDIIVETATAYIVRHGDFHYIPKSNQIG-----OPTLPNNSLATPS 536
QY 227 PSLPNTGTSHEKHEEDGYDFDANRIIAEDSGFVMSHGDNHNYFFFKDLTEEQIKAQK 286
DB 537 PSLPNTGTSHEKHEEDGYDFDANRIIAEDSGFVMSHGDNHNYFFFKDLTEEQIKAQK 596
QY 287 HLEEVKTSRSHGLDSSSHEQDYPSNAKEMKDLKXIEKIAKINQYGVKRESIVVWKEK 346
DB 597 HLEEVKTSRSHGLDSSSHEQDYPSNAKEMKDLKXIEKIAKINQYGVKRESIVVWKEK 656
QY 347 NAIIVPHGDHHDADIDEPKPVGIGHSNYSNELFKPESGVAKKGNKYITGEELTNVYNL 406
DB 657 NAIIVPHGDHHDADIDEPKPVGIGHSNYSNELFKPESGVAKKGNKYITGEELTNVYNL 716
QY 407 LKNSFNQNFNLANGKRVSFSPPELEKLGINKLVKITPDGKLVLEKYGKVGEGV 466
DB 717 LKNSFNQNFNLANGKRVSFSPPELEKLGINKLVKITPDGKLVLEKYGKVGEGV 776
QY 467 GNIAFELDQPLPQGTTFKYTIASKDYPEVSDGTFVPTSLAYKMASQTIFFPFHAGDT 526
DB 777 GNIAFELDQPLPQGTTFKYTIASKDYPEVSDGTFVPTSLAYKMASQTIFFPFHAGDT 836
QY 527 YLRVNPQFAVPKGTDALVRVDFEFGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPV 586
DB 837 YLRVNPQFAVPKGTDALVRVDFEFGNAYLENNYKVGEEKLPKLNQGTTRTAGNKIPV 896
QY 587 TPNANAYLDNQSTYIVVEPILKEKQTDKPSILPOFKKNKAQENSKLDEKVEEPTSEKV 646
DB 897 TPNANAYLDNQSTYIVVEPILKEKQTDKPSILPOFKKNKAQENSKLDEKVEEPTSEKV 956
QY 647 EKEKLETSNNTSLTEVPTVDPQEKVAFESYGMKLENVLFPNMDGTIELYLPDGE 706
DB 957 EKEKLETSNNTSLTEVPTVDPQEKVAFESYGMKLENVLFPNMDGTIELYLPDGE 1016
QY 707 VIKKNMADFTGAPQNGENKPSNGKYSTGTGVENQPTENKPADSLPEAPKPVKPPNS 766
DB 1017 VIKKNMADFTGAPQNGENKPSNGKYSTGTGVENQPTENKPADSLPEAPKPVKPPNS 1076
QY 767 TONGMLNPGNVGSDPMLDSALEAPADVPQEKLEKTASYGLDLSVIFNMDGTIELR 826
DB 1077 TONGMLNPGNVGSDPMLDSALEAPADVPQEKLEKTASYGLDLSVIFNMDGTIELR 1136
QY 827 LPSGVEIKKNL 837
DB 1137 LPSGVEIKKNL 1147

RESULT 15

AAU84056

ID AAU84056 standard; Peptide: 1238 AA.

XX AC AAU84056;

XX DT 08-MAY-2002 (first entry)

XX DE S. pneumoniae derived chimeric peptide, NEW29.

XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX KW pneumonia; streptococcal bacterial infection; mutant; mutin;

XX KW BVH-11-2.

XX OS Streptococcus pneumoniae.

XX OS Synthetic.

XX PN WC2001198334-A2.

XX PD 27-DEC-2001.

XX XX 19-JUN-2001; 2001WO-CA00908.

XX XX 20-JUN-2000; 2000US-212683P.

XX XX (SHIR-) SHIRE BIOCHEM INC.

XX PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

XX XX WPI; 2002-122272/16.

XX PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX PT epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia

XX PS Example 1; Page -; il3pp; English.

XX CC The invention describes an isolated polypeptide (I) with 70-90%
XX CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX CC comprising (I) is useful for therapeutic or prophylactic treatment of
XX CC meningitis, otitis media, bacteraemia or pneumonia infection in an
XX CC individual susceptible to these disorders. (II) is also useful for
XX CC therapeutic or prophylactic treatment of any streptococcal bacterial
XX CC infection (e.g., caused by Streptococcus pneumoniae, group A
XX CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
XX CC Staphylococcus aureus) in an individual susceptible to the infection.
XX CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX CC techniques. The Streptococcus polypeptides are useful in a diagnostic
XX CC test for S. pneumoniae infection. (III) is useful for designing DNA
XX CC probes for use in detecting the presence of Streptococcus in a biological
XX CC sample suspected of containing the bacteria. The DNA probes may also be
XX CC used for detecting circulating S. pneumonia nucleic acid in a sample for
XX CC diagnosing streptococcal infections. This sequence represents a chimeric
XX CC gene created from fragments and variant fragments of Streptococcus
XX CC pneumoniae genes, described in the method of the invention.
XX CC Note: This sequence does not appear in the specification but has
XX CC been created according to information given in the invention.

XX SQ Sequence 1238 AA;

Query Match 77.4%; Score 3402.5; DB 23; Length 1238;

Best Local Similarity 79.6%; Pred. No. 1.9e-200;

Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11;

QY 10 SQENKDNRVSYVDSQS--SQKSENL-----TPDQVSKQEGIQAEQI-VIKIT 55

DB 441 SSTAASDNTQSVAKGTSKPAKNSLSLKLKELYDSPAQRYSDESLGDFDPAKIIISRT 500

Qy	56	DQGYVTSHGHDHYHYNKGVYPYDALFS-BELLMKDPNYOLKDADLVNEVGKYI IKVDGKY	114
		I : : : : : I : : : : : I : : : : :	
Db	501	PNGVAIPHGDHYHF---IPYSKLSALEPKEAR-----MVPISGT- 536	
Qy	115	VYVLKDAAHADNVRKDEINRQKHVDKNFKSVNAVARSQRYTT-----NDGYV	167
Db	537	-----GSTVST-----NAKPNEVVSLGSSNPSSLITSKELSSASDGYI	577
Qy	168	FNPADIETGNAYIVPHGGHYHYIPKSDLSASELAAKAHLAKGNQPQLSYS-STPS	226
		II : : : : II : : : : III : : : : II : : : : II : : : :	
Db	578	FNPKDIVEETFATAYIVRHGDHEHYIPKSNIQIG-----OPTLPNNLSLATPS	622
Qy	227	PSLPINPGTSHKEHEEDGCGFDANRIIAEDSESGFVMSHGDNHHFFPKDLTEQIQRAAQK	286
		II : : : : II : : : : II : : : : II : : : : II : : : :	
Db	623	PSLPINPGTSHKEHEEDGCGFDANRIIAEDSESGFVMSHGDNHHFFPKDLTEQIQRAAQK	682
Qy	287	HLEVKTSHNLDSLSSHODPYSPNAKEMKLDKKIEBK IAGIMKOYGVARRESIVVNKEK	346
Db	683	HLEVKTSHNLDSLSSHODPYSPNAKEMKLDKKIEBK IAGIMKOYGVARRESIVVNKEK	742
Qy	347	NALIYPHGDHHHADPIDEHKPVGIGHSHSNVELFKPEEGVAKKEGNKVYIGEELTNVVNL	406
Db	743	NALIYPHGDHHHADPIDEHKPVGIGHSHSNVELFKPEEGVAKKEGNKVYIGEELTNVVNL	802
Qy	407	LKNSTENNQNTLANGOKVSFSFPPELEKLGINMLVKLLITDGK VLEKVGSKVGEV	466
Db	803	LKNSTENNQNTLANGOKVSFSFPPELEKLGINMLVKLLITDGK VLEKVGSKVGEV	862
Qy	467	GNTANFELDOPYLPQGTFKYTTASKDYPEVSDGTFVTPTSLAYKMASQITFYPFHAGDT	526
Db	863	GNTANFELDOPYLPQGTFKYTTASKDYPEVSDGTFVTPTSLAYKMASQITFYPFHAGDT	922
Qy	527	YLRVNPQFAVPKGTDALYRVDFEFGHNAYLENNYKVGIEIKLP IPLKNOGTRTAGNKIPV	586
Db	923	YLRVNPQFAVPKGTDALYRVDFEFGHNAYLENNYKVGIEIKLP IPLKNOGTRTAGNKIPV	982
Qy	587	TFWANAYLONQSYIIVEVDPILEKENOTDKPSTLPQFKRNAQAENS K LDEKVEBPKTSEKV	646
Db	983	TFWANAYLONQSYIIVEVDPILEKENOTDKPSTLPQFKRNAQAENS K LDEKVEBPKTSEKV	1042
Qy	647	EKELSETGNSTNSNLTLEEVPVDVQEKVAKFAESYGMKLENVLFNM DGTIELYLPSPGE	706
Db	1043	EKELSETGNSTNSNLTLEEVPVDVQEKVAKFAESYGMKLENVLFNM DGTIELYLPSPGE	1102
Qy	707	VIKKNADFTGEAPQNGENKPSENGKYSTGTVENQPTENKPADSL PEA NEXPKVPKENS	766
Db	1103	VIKKNADFTGEAPQNGENKPSENGKYSTGTVENQPTENKPADSL PEA NEXPKVPKENS	1162
Qy	767	TDNGMLNPEGNVGSDPMILDSALEEPAVDVQEKLEKFTASYGL GDSVFIFNMDGTIELR	826
Db	1163	TDNGMLNPEGNVGSDPMILDSALEEPAVDVQEKLEKFTASYGL GDSVFIFNMDGTIELR	1222
Qy	827	LPSGEVIKKNL 837	
Db	1223	LPSGEVIKKNL 1233	

Search completed: May 13, 2003, 13:53:44
Job time : 45.8944 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: May 13, 2003, 13:51:27 : Search time 19.0029 Seconds

(without alignments)
4249.498 Million cell updates/sec

Title: US-09-471-255-16

Perfect score: 4396

Sequence: 1 CAVALNQHRSQENKDNRRVS.....GTIELRPSGEVKKNLLIS 840

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-73.*

1: piri.*

2: piri.*

3: piri.*

4: piri.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4274.5	97.2	1039	2 H95115	conserved hypotet
2	4264.5	97.0	1039	2 D97985	hypothetical prote
3	1015.5	23.1	855	2 D98004	histidine Motif-co
4	1008.5	22.9	853	2 C97985	hypothetical prote
5	989	22.5	839	2 B95136	conserved domain p
6	987	22.5	839	2 G95115	conserved domain p
7	981.5	22.3	828	2 E98004	hypothetical prote
8	975.5	22.2	802	2 C95136	conserved domain p
9	895.5	19.5	822	2 T46758	hypothetical 92.4K
10	313.5	7.1	182	2 F97985	hypothetical prote
11	212.5	4.8	961	2 G90053	hypothetical prote
12	210.5	4.8	2004	2 F95133	immunoglobulin Al
13	196.5	4.5	940	2 B97002	fibronectin-bindin
14	194.5	4.4	1963	2 S19002	IgA-specific metal
15	194.5	4.4	2485	1 H71621	serine/threonine-s
16	191.5	4.4	1939	2 T18372	repeat organellar
17	189	4.3	2269	2 T28677	rhodopy protein -
18	187	4.3	1873	2 T30944	rhodopy protein pr
19	186.5	4.2	1849	2 C41859	IgA-specific metal
20	180	4.1	1231	2 T28681	rib protein, Stre
21	180	4.1	1435	2 A37793	erythrocyte-bindin
22	179	4.1	1650	2 T18444	hypothetical prote
23	177.5	4.0	1271	2 A45555	glutamate rich pro
24	176	4.0	2481	2 D90011	FmtB protein (impo
25	175.5	4.0	940	2 AB1744	internalin protein
26	175.5	4.0	1134	2 A60234	IgA Fc receptor pr
27	175.5	4.0	1164	1 FCSOAG	IgA Fc receptor pr
28	174	4.0	1139	1 E64234	cytadherence-acces
29	173.5	3.9	786	2 T18469	hypothetical prote

RESULT 1

H95115

conserved hypothetical protein SP1004 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: H95115

R:Tetrelin, H.J.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl
nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: H95115

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1039 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:gl4972477; GSPDB:GN00164; TIGR:

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP1004

Query Match 97.2%; Score 4274.5; DB 2; Length 1039;

Best Local Similarity 82.3%; Pred No. 4.2e-209;

Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAVALNQHRSQENKDNRRVS YVDGSSQSKSENLTDPQVSQKEGIAEQIIVIKITDQGYV 60

DB 21 CAVALNQHRSQENKDNRRVS YVDGSSQSKSENLTDPQVSQKEGIAEQIIVIKITDQGYV 80

QY 61 TSHGDHYHYNGKVPYDALFSELLMKDPNYQLKDADIVNEKGGYIIKVDGKYYVYLDK 120

DB 81 TSHGDHYHYNGKVPYDALFSELLMKDPNYQLKDADIVNEKGGYIIKVDGKYYVYLDK 140

QY 121 AAHADNVRTKDEINRQKHVDKNEKVNNAVARSQGRYTNDGYVFNPAIDTGN 180

DB 141 AAHADNVRTKDEINRQKHVDKNEKVNNAVARSQGRYTNDGYVFNPAIDTGN 200

QY 181 YIVPHGGHYHYIPKSDLSASELAALAAKAGLAGNMQPSQLSYSTASDNNVTQVAKGSTSK 260

DB 201 YIVPHGGHYHYIPKSDLSASELAALAAKAGLAGNMQPSQLSYSTASDNNVTQVAKGSTSK 260

QY 224 PANKSENLOSLLKELYDPSAQRYSSESDGLVDFPAKIIISRTPNGVAIPHGDHYHFTPYSK 320

DB 261 PANKSENLOSLLKELYDPSAQRYSSESDGLVDFPAKIIISRTPNGVAIPHGDHYHFTPYSK 320

QY 224 LSALAEKIAMWPISCTGSTVSTNAKPNVWSSLSLSPSSLATSKELSSASDGYIFN 380

DB 321 LSALAEKIAMWPISCTGSTVSTNAKPNVWSSLSLSPSSLATSKELSSASDGYIFN 380

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

QY 224 TPSPSLPINPGTSHKHEED 243

DB 224 TPSPSLPINPGTSHKHEED 243

```
Db 381 PKDIVETATAYIVRHGDHFIYIPKSNQIGOPTLPNNSLATPSPSLPINPOTSHKHEED 440
Qy 244 GYGFANDRIIAEDSESGFYNSHGDHNFYFFKDLTTEQIKAAQKHEEVKTSNGLDLSLS 303
Db 441 GYGFANDRIIAEDSESGFYNSHGDHNFYFFKDLTTEQIKAAQKHEEVKTSNGLDLSLS 500
Qy 304 HQDYPNNAKEMKDLKKIEKIAKIMQYGVKRESIVVNKKNALIIYPHGDHHAADPID 363
Db 501 HQDYPNNAKEMKDLKKIEKIAKIMQYGVKRESIVVNKKNALIIYPHGDHHAADPID 560
Qy 364 EHKPVGIGSHSNYELPPEEGVAKKEGKNYVTGEBELTNVNNLLKNSFTNNQNTLANGQ 423
Db 561 EHKPVGIGSHSNYELPPEEGVAKKEGKNYVTGEBELTNVNNLLKNSFTNNQNTLANGQ 620
Qy 424 KRVSFPPELEKKGILGNMLVKLITPDGKVKLEKSGKVGEGVGNIANFELDQYLPQGT 483
Db 621 KRVSFPPELEKKGILGNMLVKLITPDGKVKLEKSGKVGEGVGNIANFELDQYLPQGT 680
Qy 484 FKYTASDKYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 543
Db 681 FKYTASDKYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 740
Qy 544 VRVDFEFGHGNAYLENNYKVGIEIKLPIPKLNQOTTRTAGNKIPVTFMANAYLDNOSTYIVE 603
Db 741 VRVDFEFGHGNAYLENNYKVGIEIKLPIPKLNQOTTRTAGNKIPVTFMANAYLDNOSTYIVE 800
Qy 604 VPILKENOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTL 663
Db 801 VPILKENOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTL 860
Qy 664 EKVPTVDPQVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVGIKKMADFTGEAPQGN 723
Db 861 EKVPTVDPQVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVGIKKMADFTGEAPQGN 920
Qy 724 GENKPSGKVGSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVSDPM 783
Db 921 GENKPSGKVGSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVSDPM 980
Qy 784 LDSALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIRKKNL 837
Db 981 LDSALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIRKKNL 1034

RESULT 2
D97985
hypothetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: D97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:AE007317; PID:AAK99712.1; PID:g15458515; GSPDB:GN00174
C:Genetics:
A:Gene: phtE

Query Match 97.0%; Score 4264.5; DB 2; Length 1039;
Best Local Similarity 82.1%; Pred. No. 1.3e-208;
Matches 832; Conservative 2; Mismatches 3; Indels 177; Gaps 1;

Qy 1 CAYALNQHRSQENKDNRRYSYVDGSSQSKSENLPDQVQSKEGIAEQAEIVIKITDQGVV 60
Db 21 CAYALNQHRSQENKDNRRYSYVDGSSQSKSENLPDQVQSKEGIAEQAEIVIKITDQGVV 80
Qy 61 TSHGDHYHYNGKVPYDALFSELLMKDPNYQLKDAIVNEVKGYYIIVKDGKYYVYVKD 120
```

```
Db 81 TSHGDHYHYNGKVPYDALFSELLMKDPNYQLKDAIVNEVKGYYIIVKDGKYYVYVKD 140
Qy 121 AAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRYTTNDGYVFNPAIIEDTGNAL 180
Db 141 AAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRYTTNDGYVFNPAIIEDTGNAL 200
Qy 181 YIVPHGHGHYHIFKSDLSASELAHAAKHAHLAGKNQPSQLSYSS----- 223
Db 201 YIVPHGHGHYHIFKSDLSASELAHAAKHAHLAGKNQPSQLSYSS----- 260
Qy 224 ----- 223
Db 261 PANKSENQLSKELYDSPAQRYSSESDGLVFDPAKIIISRTPNGVAIPHGDHYHIFPYSK 320
Qy 224 ----- 223
Db 321 LSALEEKIARVPISGTGSTVSTNAKPNNEVSSGLSSNPNSSLTTSKELSSASDGYIFN 380
Qy 224 -----TPSPSLPINPGTSHKHEED 243
Db 381 PKDIVETATAYIVRHGDHFIYIPKSNQIGOPTLPNNSLATPSPSLPINPOTSHKHEED 440
Qy 244 GYGFANDRIIAEDSESGFYNSHGDHNFYFFKDLTTEQIKAAQKHEEVKTSNGLDLSLS 303
Db 441 GYGFANDRIIAEDSESGFYNSHGDHNFYFFKDLTTEQIKAAQKHEEVKTSNGLDLSLS 500
Qy 304 HQDYPNNAKEMKDLKKIEKIAKIMQYGVKRESIVVNKKNALIIYPHGDHHAADPID 363
Db 501 HQDYPNNAKEMKDLKKIEKIAKIMQYGVKRESIVVNKKNALIIYPHGDHHAADPID 560
Qy 364 EHKPVGIGSHSNYELPPEEGVAKKEGKNYVTGEBELTNVNNLLKNSFTNNQNTLANGQ 423
Db 561 EHKPVGIGSHSNYELPPEEGVAKKEGKNYVTGEBELTNVNNLLKNSFTNNQNTLANGQ 620
Qy 424 KRVSFPPELEKKGILGNMLVKLITPDGKVKLEKSGKVGEGVGNIANFELDQYLPQGT 483
Db 621 KRVSFPPELEKKGILGNMLVKLITPDGKVKLEKSGKVGEGVGNIANFELDQYLPQGT 680
Qy 484 FKYTASDKYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 543
Db 681 FKYTASDKYPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 740
Qy 544 VRVDFEFGHGNAYLENNYKVGIEIKLPIPKLNQOTTRTAGNKIPVTFMANAYLDNOSTYIVE 603
Db 741 VRVDFEFGHGNAYLENNYKVGIEIKLPIPKLNQOTTRTAGNKIPVTFMANAYLDNOSTYIVE 800
Qy 604 VPILKENOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTL 663
Db 801 VPILKENOTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTL 860
Qy 664 EKVPTVDPQVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVGIKKMADFTGEAPQGN 723
Db 861 EKVPTVDPQVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEVGIKKMADFTGEAPQGN 920
Qy 724 GENKPSGKVGSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVSDPM 783
Db 921 GENKPSGKVGSTGTVENQPTENKPADSLPEAPNEKVPKPNSTDNGLNPEGNVSDPM 980
Qy 784 LDSALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIRKKNL 837
Db 981 LDSALEAPAVDPVQEKLEKFTASYGLDLSVIFNMDGTIELRLPSGEVIRKKNL 1034

RESULT 3
D98004
histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
```

J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99864.1; PID:gl5458682; GSPDB:GN00174
C:Genetics:
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 23.1%; Score 1015.5; DB 2; Length 855;
Best Local Similarity 30.3%; Pred. No. 6.5e-44;
Matches 281; Conservative 136; Mismatches 232; Indels 277; Gaps 31;

QY 1 CAYALNQRSQENKNNRVSYVDGSSQSKSENLPDQVSQKEGQAQIVIKITDQYV 60
DB 31 CAYELGHOAQVTKNNRVSYIDGQATQKTENLPDEVSKREGNAQIVIKITDQYV 90
QY 61 TSHGDHYHYNGKVPYDALFSELLMKDPNYOLKADIVNEVKGYYIIKVDGKYYVYLK 120
DB 91 TSHGDHYHYNGKVPYDAIISELLMKDPNYOLKDEDIISEIKGGYIVKVDGKYYVYLK 150
QY 121 AAHADNVRTKDEINROKQEHVXDNE-----KVNNSNVAARVQGRYTTNDGYVFNPAIDIED 176
DB 151 AAHADNVRTKEINROKQEHVXDNE-----KVNNSNVAARVQGRYTTNDGYVFNPAIDIED 210
QY 177 TGNAYTVPHGGHYHYIPKSDLSASELAQAHAHLAKGNQPSOLSYSSNPS-----P 227
DB 211 TGDAYTVPHGGHYHYIPKSELASELAQAHAHLAKGNQPSOLSYSSNPS-----P 270
QY 228 SLIPINPGTSHXHEEDGYGFANDRIIAEDSGFVMSHGDHNNHYFFKDLTSEIQIAAKQH 287
DB 271 SVS-NGPTTN----- 279
QY 288 LEEVKTSHNGLDSSHEQDYPNAKEMKDLKKTIEKTAGIMKOYGVKRESI-----V 341
DB 280 ---TNTSN-----SNTSQASQNDISLLKOLYKLPDSRHVESDGLIFDPAQI 327
QY 342 VRKEKNAIYTPHGHADIPDEHPVPVIGHSHSNYELFKPEGVAKKEGNKVYVGEELT 401
DB 328 TSTANGVAVPHGDHYHFIPYLSPL-----EELKLA 359
QY 402 NVNLLKNTFNQNTFLANGKRVVSFPPE-----LEKKLGINWL 443
DB 360 RIPL-----RYRSNHWVPDSRPEQSPQSTPEPSPQAPNPQAPNPIDEKL-VKEA 414
QY 444 VKLITPDGKVLKGVKGVFEGVGNIANFELDQVLPQGTETKTIASKDYPEVSYDGTFT 503
DB 415 VRKV-CDGYVFE-----NGVPR-----YIPAKDLSAETA 451
QY 504 VPTSLAYKASOTIFYPFHAGDTYLRVNFQFAPVPGTDLALRVFDEFGNAYLENNYKV 563
DB 452 KOESLSHKLGAKTDLDPSSDREFY-----NKAYDILLARIHQDLDLKNQGRVDFA 501
QY 564 ETKLPKLNQGTTRTAGNKIPVTFEMANAYLDNQSYIIVEVILEKENGTDKPSILPOFK 623
DB 502 -----LDNLLERLKVSSDKVLVDILAFL-----APIRPE-RLGPNQAIIY 546
QY 624 RNKAQENSL-----DEKVEEPK--TSEK-----VEREKLSETGNSTSN 660
DB 547 DBEIQV-AKLAGKTYTDCGYIFDRDITSDGDAYVTPHMTSHWIKKDLSEASRAAAQ 605
QY 661 STLEEYPTVDPQE-----KVAKFAESY-----GKLENLVFNMDGTIELYLP 706
DB 606 AVAKEGLTPPSTDHODSGNTAKGAALYNRVKAARKVPLDRMPYNLTQYVE--VKRGS 663
QY 707 VIKKNMADFTG-----EAPQG-----NGENKPSD- 730
DB 664 LIIPHVDHYHNKFEWDEGLVEAPKGSLEDLLATVYVVEHPNPHSDNGFGNASDH 723

QY 731 -----NGKSVGTGVNPTENKPADSLPEAPNEKVPKPNSTNDGMLNPGNVSQDPM 785
DB 724 VQRKNQADINTQTE-KPNEKPOTEPKPEETPREEPQSE-----KPESPXPTPEEP 776
QY 786 SALEEAP-----AYDPVQOEKLEK 803
DB 777 ESPEESPESEPOVETEKVKEKLE 802

RESULT 4
C97985
Hypothetical protein phtD [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: C97985
C:Authors: J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; Mehren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: C97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99711.1; PID:gl5458514; GSPDB:GN00174
C:Genetics:
C:Gene: phtD
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 22.9%; Score 1008.5; DB 2; Length 853;
Best Local Similarity 31.9%; Pred. No. 1.5e-43;
Matches 295; Conservative 113; Mismatches 263; Indels 255; Gaps 31;

QY 1 CAYALNQRS--QENKNNRVSYVDGSSQSKSENLPDQVSQKEGQAQIVIKITDQY 59
DB 20 CSYELGRHQAGOVKKENRVSYIDGQAGAKAENLPDEVSKREGINAEQIVIKITDQY 79
QY 60 VTSHGDSHYHYNGKVPYDALFSELLMKDPNYOLKADIVNEVKGYYIIKVDGKYYVYLK 119
DB 80 VTSHGDSHYHYNGKVPYDAIISELLMKDPNYOLKDSIVNEIKGGYIVKVDGKYYVYLK 139
QY 120 DAHADNVRTKDEINROKQEHVXD-----NEKVNNSNVAARVQGRYTTNDGYVFNPAIDIEDTG 178
DB 140 DAHADNVRTKEEIKRQKQERSHNSRADNAAARAGQRYTTDDGYIFNASLIEDTG 199
QY 179 NAYTVPHGGHYHYIPKSDLSASELAQAHAHLAKG-----NMQPSQ-----LSYSS 223
DB 200 DAYIVPHGDHYHYIPKSDLSASELAQAHAHLAKG-----NMQPSQ-----LSYSS 259
QY 224 TSPSLPINPGTS-----HEKH-EEDGYGFANDRIIAEDSGFVMSHGHQNH 269
DB 250 TVTPTYHONQGENISSLLRELYAKPLSERHVESDGLIFDPAQITSRANGVAVPHGDHYH 319
QY 270 YFFKDKLPEEQIKAAQKHLSEVKTSHNGLDL-----SSHEQDQYPSN 311
DB 320 FIPYSLSPLEKLARIIPLYRSNHWVPDSRPEQSPQSTPEPSPQAPNPQAPNP 379
QY 312 AKEMKDKLKEETAG--IMKQGVKRESIVNAKEKNALIIYPHGDDHHADIPDEHPV 369
DB 380 PIDEK-LVKEARVKVGDGYVFEENGVP-----YIPAKDLSAETAAG 420
QY 370 IGHSHSNELFKPEGVAKKE-----GNKV-----TGEELTNV 403
DB 421 IDSKLAKQESLSHLKAGAKKTOLPSSDREFYNKAYDILLARIHQDLDLKNQGRVDFA 480
QY 404 VNLKNTFNQNTFLANGKRVVSFPPELEKLGINMLVKLITPDGKYLEKYSKVFG 463
DB 481 LERLKVSSDKVLV-----DQILAFAPIRHPERLGRPNQAIIYTTDDEIQVAKLAGYTT 536
QY 464 E-----GVCNITANFELDQVYLPQGTETKTIASKDYPEVSYDGTFTVTSLAYKASQTIF 518

Db 537 EDGYIFDPDITSDGDAYVTPHMTSHWIKKDSLSE-----KTSNGLDSL-----SSHEQDPSN 311
 Qy 519 YPFHAGTILRVNPOFAVPGKGTDALVRVDFEFGNAYLNNYKVGKIKLPKLNQGTTR 578
 Db 582 YAKEKGLT-----PPSTDHDSGNTAKGAEIYNRVKAAK-KVPL-----YIPAKDLSAATAAG 414
 Qy 579 TAGNKIPVTIMANAYLDNGSTIVIVEPILEKENQTDKPSILQFRRNKAQENSKLDEKV- 637
 Db 622 ---DRMPYNLQ-----YTVVEV-----GNKXV-----TCEELTNV 403
 Qy 638 BEPKTSEKVEKEKLSTGNTSNTSLSEVPTVDPVQEVAKTAESYGMKLENLFNMGT 697
 Db 661 EAPK-----GYSLDLL-----AT 674
 Qy 698 IEYLPSGEVIRKKNMADFTGEAPQ-----NGEN--KPSNGKVTGTGVTNOPTENKPAD 750
 Db 675 VKYV-----EHPNERPHSDNGFGNASDHVQRNKGQADTNOTE-KPNEEKQOT 722
 Qy 751 SLPEANE-----KPVKPNSTDNGLNPEGNVGSDDPMDLSALEAPAVDPVQEKLEKFT 805
 Db 723 EKPEDEKHEVSEPHPSD-----EXENHVGLNPSADNLYKPSTDETEEEAE--- 773
 Qy 806 ASYGLGLDSVIFNMGTIELRLPSGE 831
 Db 774 -----DITDEAEIPOVE 785

RESULT 5
 B95136
 Conserved domain protein Sp1174 [Imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: B95136
 R:Teitelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hoon, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, J.D.; Hickey, E.K.; Holt, I.E.
 A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: B95136
 A:Status: preliminary
 A:Molecule type: DNA
 A:Cross-references: GB:AE005672; PID:AAK75283.1; PID:g14972654; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 22.5%; Score 989; DB 2; Length 819;
 Best Local Similarity 31.9%; Pred. No. 1.3e-42;
 Matches 298; Conservative 118; Mismatches 265; Indels 252; Gaps 36;

Qy 1 CAVALNQRS-QNKNNRVSYVDSQSKSENTPTDOVSQKEIQAPQIVIKITDQGY 59
 Db 20 CSVELGRYAGQDKESNRVAVIDGDAQKAENTPDEVSKREGINAEQIVIKITDQGY 79
 Qy 60 VTSHGDRHYHYNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYIKVDGKYYVYLK 119
 Db 80 VTSHGDRHYHYNGKVPYDAIIEELMKDPNYQLKDSDIVNEIKGVYLVKNGKYYVYLK 139
 Qy 120 DAAHADNRTKDEINRQKEHYKD-NEKYNVSNVAVARSQRTYNDGVYFNPADIETDG 178
 Db 140 DAAHADNRTKEIKRQKQERSHNHNSRADNAVAAARAGRYTTDDGYTFNASDIETDG 199
 Qy 179 NAYIVPHGDRHYHYIPKSDLSASLAAKALHAGK--NMQPSOLSYSTPSP-----SL 229
 Db 200 DAIIVPHGDRHYHYIPKNSLSASLAAAEAYWNGKQSRPSSSSSYNANPAQPLSENHL 259
 Qy 230 PINPGTSH-----EKH-EDGYGFDANRIIAEDSGFVMSHGDN 268
 Db 260 TVTP-TYHQNGENISSLLRELYAKPLSERHVESDGLIFDPAQITSRARGVAVPHGNHY 318

Qy 269 HYFPKDLTTEQIKAAQKHLEEV-----KTSNGLDSL-----SSHEQDPSN 311
 Db 319 HF-----IPYEQNSELEKRIARIIPLYRSNHVVPDSRPEEPSQPTPEPSPQAPS 373
 Qy 312 AKEMKDLKKIEEKIAG--IMQYGVKRESIVVKNENAIYIYHGDHHDADPIDEKHPVG 369
 Db 374 PIDCK-LVKEAVRKVGDDGYVEENGVS-----YIPAKDLSAATAAG 414
 Qy 370 IGHSNVELFKPEGVAKKE-----GNKXV-----TCEELTNV 403
 Db 415 IDSLAKQBSLSHLKTKYDLSPSSDREFYNKAYDLLARIHQDLLDNKQVDFEALDNL 474
 Qy 404 VNLKNTFNQNTLANGQKRVSFSPPELEKKLGIMLVKLITPDCKVLEKSVKVP 463
 Db 475 LERLKDVSSDKVKLV-----EDILAFAPISHPERLGPNAQITVTDDEIQVAKLAGKTT 530
 Qy 464 E-----GVGNIANFELDPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMAQTIF 518
 Db 531 EDGYIFDPDITSDGDAYVTPHMTSHWIKKDSLSE-----AERAAQA-- 575
 Qy 519 YPFHAGTILRVNPOFAVPGKGTDALVRVDFEFGNAYLNNYKVGKIKLPKLNQGTTR 578
 Db 576 YAKEKGLT-----PPSTDHDSGNTAKGAEIYNRVKAAK-KVPL----- 615
 Qy 579 TAGNKIPVTIMANAYLDNGSTIVIVEPILEKENQTDKPSILQFRRNKAQENSKLDEKV- 637
 Db 616 ---DRMPYNLQ-----YTVVEV-----GNKXV-----TCEELTNV 403
 Qy 638 BEPKTSEKVEKEKLSTGNTSNTSLSEVPTVDPVQEVAKTAESYGMKLENLFNMGT 697
 Db 655 EAPK-----GYTLELL-----AT 668
 Qy 698 IEYL-----PSGEVIRKKNMADFTGEAPQ-----NGENKPSNGKVTGTGVTNOPTENK 747
 Db 669 VKYVVEHPNERPHSDNGFGNASDHVQRNKGQADTNOTEKPSPEEKQOTPEETPREK 728
 Qy 748 PADSLPAPNPKPY-KPENSTDNGLNPEGNVGSDDPMDLSALEAPAVDPVQEKLEKFTA 806
 Db 729 PQSEKPSP--KPTEEPESPPE-----SEPOVET-----EKVEEKLEAEED 769
 Qy 807 SYGLGLDSVI-FNMDGTIELRLPSGEVIRKKNLL 838
 Db 770 LLGKIQDPIKNSNAKETL-----TG--LKNLL 795

RESULT 6
 G95115
 Conserved hypothetical protein Sp1003 [Imported] - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: G95115
 R:Teitelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hoon, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, J.D.; Hickey, E.K.; Holt, I.E.
 A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: G95115
 A:Status: preliminary
 A:Molecule type: DNA
 A:Cross-references: GB:AE005672; PID:AAK75120.1; PID:g14972476; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 22.5%; Score 987; DB 2; Length 839;
 Best Local Similarity 32.0%; Pred. No. 1.8e-42;
 Matches 299; Conservative 106; Mismatches 278; Indels 252; Gaps 33;

	Query Match	22.38;	Score	981.5;	DB	2;	Length	828;
	Best Local Similarity	33.78;	Pred.	No.	3.3e-42;			
	Matches	272;	Conservative	102;	Mismatches	199;	Indels	233;
	Gaps							
Qy	1	CAYALNHRSENKDNRRVSYDGSQSQSKSENLTPQVSQKEGTAQBQIVIKITDQGYV	60					
Db	32	CSYELGLQARTVKENRNVSDIGKQATKTENLTPEVSKREGINABQIVIKITDQGYV	91					
Qy	61	TSHGDHYHYNCKVPDYDALPSELLMKDPVOLKADDAIVNEKVGYITKVCGKYVYLKD	120					
Db	92	TSHGDHYHYNCKVPDYDAIFSEELMKDPNKLDDEDIVNEKVGYITKVCGKYVYLKD	151					
Qy	121	AAHADNVRTKDEINRQKHQEKYONE---KYSNVAVARSGRYTTNDGYVFNPADIIED	176					
Db	152	AAHADNVRTKEIIRHQEHQESHQHBGGTPRNDGAVALARSGRYTTDCGYTFNASDIIED	211					
Qy	177	TGNAYIVPHGGHYHIIPKSLDSASLASAAKAHLAKNMQPQLSYSSTPS-----P	227					
Db	212	TGDAYIVPHGGHYHIPKNELASLASAAAAEFLSGRGLNSRTRYRONSONTSRTNNVP	271					
Qy	228	SLPINPTGSH-----EXH-EDDGXGFUDANRIIA	254					
Db	272	SVS-NPGTNTNTSNNSTNSQASQSNIDISLLKOLYKLPLSQRHVESDGLVFPQAITS	330					
Qy	255	EDESOFVMHGHDHNHFYFKDLTBEQIKAAGKHLEEVKTSINGLSDLSSSHQQDYP----	309					
Db	331	RTARGVAVPHGDHYHFTPYSQMSELEERARIILRYBSNHWWPDSRPQEQSPQOPTPEPS	390					
Qy	310	-----SNAKEMDKDKIEEKIAGIMQOYGVR-----ESIUVNKEKNAI	349					
Db	391	PGOPAPNLKIDSNSLVSQLRVAGE--GVYFEKGISRIFYAFKADLPSETVTKLESKL-	447					
Qy	350	IYPGDHHHADPIDEBKPVGIGHSHNYELEKPEBGVAKEGKNKYTGELTNVVNLL--	407					
Db	448	-----SKOESYSHILT-----AKKENVAPRODFEYDKAYNLLTE	481					
Qy	408	-----KNSYF-----NNONFTLANGQRV-----SFSPPELEKKLGINNLYKL	446					
Db	482	AHKALFENKNGNSDFOALDKLERLNDESTNKEKLVDDLLAFLAPIHPERUG-----	534					
Qy	447	ITPGKYLEKSYGVFGVGVCNIANFELDOPYLPQTPKYTIAS----KDXPEVSYDGTFF	502					
Db	535	-KPNOSI-----EYTEDVRIAQLADKYITSGYITDEHDIISDESDA	576					
Qy	503	TVPTSLAY-----KMASQITFYPPFHAGDTYLRVNPOFAVFKGTDALVRVEDE	549					
Db	577	VYTFHMGHSHWIGKDSLSDKEKVAQA--YTKEG-----ILPPSPDADVKANPT	624					
Qy	550	FHGNAYLENKYKGEIKLPIPKLNOGTRTA-----GNKI-----PVTMANAYLDNQ	597					
Db	625	GDSAAAYNRVK-GEKRLPLRLPYMVHEHTVEVRNGNLIIPH KDHYHNKIF---AWFDH	680					
Qy	598	S-----TYIVEVP-----ILEKNOTDKPSILPQEKRN	625					
Db	681	TYKAPNGYTLDELATIKYYVHEHDERPHSNDGWGNASEHVGLGKHDSEDN-----KNF	735					
Qy	626	KAQENSKLDEKVEEPKSEKYEKEL	651					
Db	736	KADE-EPVEETPAPEVQ-VETEKV	759					

Db 517 DY-----FYARYQLDLTQI--AFAEQELMLKDKKHRYD-----IVDTGIEPRLAVDVSSLP 566
Qy 521 FHAGDTYLRVNPQFAVPGKTDALVRVFEDEHGNAY--LENNYKGEIKLPIPKLNQGTTR 578
Db 567 MHAGNATYDTGSSVIPH-----IDHIHVYPYSLTRN-QIATIK----- 605
Qy 579 TAGNKKIPVTFMANAYLDNQSTYIVVPP-----ILEKNQTDKPSILQFQRNKAQNSKL 633
Db 606 -----YNNQHPVPRDVMWSKPGHEESGVPINV-----TPL 636
Qy 634 DEKVEEPK-----TSEKVEKEKLETSNSTLEEVPTVDPQVKVAFKFAESYGMKLE 688
Db 637 DKRAGMPWQIHHAEVQK-ALAEGRFAADGYI-----FDP-RDVLAQ-EFVWK-- 685
Qy 689 NYLFNMDGTIELYLPFSGEVIKK-NMADFTGEAPOGNGENKPSKGVSTGTGVNQPTENK 747
Db 686 -----DGSFIPRADGSSRLTRINKSLSQAEWQQAQELLAKKNAGDAITD--DKPEEK 737
Qy 748 PADSLPEAPNEKP---VKPENSTONGMLNPGNVGSDPMLDSALEEPAFVDPQVKLEKF 804
Db 738 QADKSNE--NOQSEASKEESDDFI-----DSLPGYGLDRATLEHDHINQLAQK- 785
Qy 805 TASYGLGLDS--VIFNMDGTIELRLPSGEVI 833
Db 786 -----ANIDPKYLIFQPEG-VQFYKNKGELV 810
RESULT 10
F97985
hypothetical protein phb-truncation [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: F97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: F97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99714.1; PID:g15458517; GSPDB:GN00174
C:Genetics:
A:Gene: phb-truncation
Query Match 7.1%; Score 313.5; DB 2; Length 182;
Best Local Similarity 39.4%; Pred. No. 2.6e-09;
Matches 76; Conservative 20; Mismatches 62; Indels 35; Gaps 6;
Qy 153 VARSGRYTNDGYVFNADITDGNAYIVPHGHHYVTPKSDLSASELAARAHLAGK 212
Db 3 VTPNKGVSNAVDGCVFNENDIVDTGDAYIVRHGDHYHVIKSSL----- 47
Qy 213 NMOPSQLSYSTPPSLPI--NPGTSHEKHEEDGYGFDANRIIAEDESFGVNSHGDHNY 270
Db 48 NNPPSHNSNTEEVGSSSSVLSNPRLVRHHEEDGCHGFANRIISEDSCGFVPHGDHNY 107
Qy 271 PFKKDI.TEQIKAAQKHLEEVKTSNGLDLSLSSHEQDQVPSNAKEMKLDKKIEKIA--- 327
Db 108 -----IKVQTKGYE--AALKNKIPSLQS-----NYPGTFDEKAVLAKVDQLLADR 152
Qy 328 GIMQYGVKREST 340
Db 153 SIYKDKPIEQRI 165
RESULT 11
G90053
hypothetical protein fnbB [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: G90053
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: G90053
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-961 <KUR>
A:Cross-references: GB:BA000018; PID:g13702452; PIDN:BAB43593.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: fnbB
Query Match 4.8%; Score 212.5; DB 2; Length 961;
Best Local Similarity 21.4%; Pred. No. 0.0042;
Matches 176; Conservative 110; Mismatches 287; Indels 251; Gaps 42;
Qy 26 QSSQKSE-NLTPQVSOKEGIAEQIVIKITDQGYVTSHG--DHYHYNGKVPYDALPSE 82
Db 290 QGSQKVEVTLQCNKVSKEFDI--KYLDGVKDRMGVTNGRIDTLNKEGK-----FSH 340
Qy 83 ELLMKDENYOLKDAIVNEVKGCVILKVGQ---KYVYVLDAAHADNVRTKDEINPKQOE 139
Db 341 FAYVKNQSLTSVTGQVTSQTKOSANNPTVKVYKHIGSDLAESVIAKLD-DTSKPE 399
Qy 140 HVKDNEKXNVNAVARSQGRYTTNDGYVFNPAADIIEDTGNAYIVPHGHHYVTPKSDLSA 199
Db 400 DV--TEKNLS-----YTSNGGYTLNLGDL--DNSKDYIVKYEGEY-----DQNA 440
Qy 200 SELAAKAHLAKGNMQPSQLSYSTSPSPSPINPGTSHKHEEEDGYGFD--NRITAEDE 257
Db 441 KDL-NFRTHUSGHHYKYPYYPYVQVLTWNGVAFYNNAGKDGCKPNOPIEKSE 499
Qy 258 SGFVMSHGDHNNHFFKKDLTEQIKAAQKH-----LEEVKTS-----HNGDLSLSHE 305
Db 500 -----PIDLDIKSEPPVEKHELTGTEIESNDSKPIDFEYHTAVEGAEGA 544
Qy 306 QDPSNAKEMKLDKKIEKIAQIMKQYGVKRESIVVYNNKNAIIPHGDDHADPIDEH 365
Db 545 E-----GIIE---TEDSIHVDFEEST---HENSKHHADVVEYE 577
Qy 366 KPVGIG---HSHSNVELFKPE--EGV-----AKKGNKYVTGELTNVNL----- 406
Db 578 EDTNFGGQVTTESNLVFEDEESTKGIIVTGAVSDDHTIVEDTKEYTTE--SNLIELVDLP 635
Qy 407 -----LKNSTFNQNT---LANGQRVSFSPFPELEKLGINMLVKLITPDGKVL 454
Db 636 EEHQGAQGPTEETENNNHHSHSGLGTENGHNHYGVIDEIEENSHVDIKSEL----- 687
Qy 455 EKVSQGVFGEVGNIAFELDQVLPQGTFTYITASKDYPEVSYDGTFTVPTSLAYKMAS 514
Db 688 -----GYEGGQNSGN-QSFEED-----TEDEKPKYEGGNI----- 717
Qy 515 QTIFFPFHAGDTYLRVNPQFAVPGKTDALVRVFEDEHGNAYLENNYKGEIKLPIPKLNQ 574
Db 718 -----VDIDF-----DSVPQIHGQNNQNSQFEDTEEDK-----PKYEQ 751
Qy 575 GTTITAGNKKIPVTFMANAYLDNQSTYIVVPILEKENQDKPSILQFQRNKAQNSKLD 634
Db 752 G-----GNIIDIDSDSVFQIHGPNKH---NEIIEEDTNKDKPNY--QF-----GHNNSVDF 797
Qy 635 EKVEEPTSEKVEKEKLETSNSTLEEVPT-----VDPVQEKVAKFAESYGMKLEN 689
Db 798 EEDTLPKVSGQNEGQQTIEDTTPPTPEVPSEPTPTPTPEVPSEGEPTPKPE- 856
Qy 690 VLFNMDGTIELYLPFSGEVIKKNMADFTGEAPQNGE-----NKPSENGKYSTG 737
Db 857 ----VPSEPTFVPP-----TPVEPSEKPPVPAKEEPKPSKPEVQEGKVITP 901

QY 738 TVE-NQ-----PTENKPA--DSLPEAPNEKPVKPNSTNDGML 772
Db 902 VIEINEKRVAVAPTQKQSKKSELPTGTEE-----STNRGML 939
RESULT 12
F95133
immunoglobulin A1 proteinase [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 14-Sep-2001
C:Accession: F95133
R:Rettellin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95133
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2004 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75263.1; PID:gl4972632; GSPDB:GN00164; TIGR:SP4
C:Genetics:
C:Gene: Spl154
C:Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase
Query Match 4.8%; Score 210.5; DB 2; Length 2004;
Best Local Similarity 22.1%; Pred No. 0.016;
Matches 178; Conservative 93; Mismatches 241; Indels 295; Gaps 47;
QY 186 GGHYHYPKSDLSASELAARAAKHAHLAKGNMQPSQLSY-----SSTP 225
Db 47 GVHYKYVADSELSSSE-----KKQLVDIPTVYVENDETYYLYVKNLSNQ 92
QY 226 SPSPLPINPSTSEKHE-EDGYGFDANRIIAEDSEGVSHGDNHYPFKKDLTEQIIAA 284
Db 93 LAELP-NTGSKNERQALVAGASLAANGILI-----FAVS-----KKRKNKTV--- 134
QY 285 QXHELEVKTSHNG-----LDLSSSH-----EODYPSNAEMKDKKIE-EKTAGIMKQY 334
Db 135 -LHLVLVAGIGNGVLSVHALENHULLNVTYDYLTSCEKPLPKREISGYTYIGYIKSK 193
QY 335 VKRESIVNKENALIIY-----HGDH-----HHADPIDEHKPVGIG--- 371
Db 194 TTSESEVSN-QKSSVATPTKQOKVDYNTPTNPDHPTVQAIQEQTPVSSTKPTVEQVVE 252
QY 372 -----HSHSNYELFKPEEGVAKKGNKVT-----GEELNV 403
Db 253 KPFSTELINPRKEEKOSDSQOLAEHKNLET-KKEKISPEKTVNLAPQDEVLSEQ 311
QY 404 VN-----LLKNSTFN-----NONFTLANGQKRVSPSPPELEKLGIMN-LVKLIT- 448
Db 312 LAKPELLYREETMETKIDPQEQIENPDLAEGTVR-----KQEGKLGKVKVEIVRIEVS 365
QY 449 -----PDGKVLKYSVGK-VFGE-----GV-----GNIANFELDQPYL 479
Db 366 NKEEVSREIVSTSTAPSPRIVEKTKTKTQVTKQEPGTVGVEHKDVQSGAIVPEAI-QPEL 424
QY 480 PGQTKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQ-----FA 535
Db 425 P-----EAVVSDKGEVQV-----TLPEAVV-----TDKGET-EVQEPSPQTV 463
QY 536 VPBKTDALVRVDFEFGNAYLENNYKVEIK--LPIKL-NQGTTRTAGNKIPVTFMANA 592
Db 464 SKQGEQVAPLPYKGN-----ISQVPTPEVTKSGQGPBKT--EEVPV----- 507
QY 593 YLDNQSTYIVPEILEKENQTDKPSLPQFRN-----KAQENSKLDKVEEPEKTSK----- 645
Db 508 -----KPEETPVNPNEGTTEGTSI--QEAENPVQPAEESTTNEKV-SPDTSKNTGE 558
QY 646 -----VEKEKLSGTGNTGNTSNTSILEEVPVTPVQEVAKFAESYGMKLENVL 691

Db 559 VSSNPSTSTSVGESNPEINDSKNSENSEKTVBEVP-VNP----- 597
QY 692 FNMGTIELYLPSEVGIKKKNMADFTGEAPGNGENKSPSENGKV---STGTVENQPTENKP 748
Db 598 --NEGTV-----GTSNQETEKPVQABETQTSNKIANENTGEVSNKPSDSKP 644
QY 749 --ADSLPEAPNEKPVKPNSENS-----TDNMLNPE---GNVSDPMLDSALEEAPVQVE 799
Db 645 PVESNAPENKNGTATKPNSENGTTSSENGQTEPEPSNGNSTEDVSTESNTSNGNEIKQ 704
QY 800 KLEKFTASYGLGLDSVIFNMDGTIELR 826
Db 705 ENE-----LDPDKKVEEPEKTELELR 724
RESULT 13
S19702
fibronectin-binding protein B - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Oct-1999
C:Accession: S19702
R:Joensson, K.; Signaes, C.; Mueller, H.P.; Lindberg, M.
Eur. J. Biochem. 202, 1041-1048, 1991
A:Title: Two different genes encode fibronectin binding proteins in Staphylococcus au
A:Reference number: S19702; MUID:92111475; PMID:1837266
A:Accession: S19702
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-940 <JOE>
A:Cross-references: EMBL:X62992; NID:949040; PIDN:CAA44726.1; PID:g581562
C:Keywords: fibronectin binding
Query Match 4.5%; Score 196.5; DB 2; Length 940;
Best Local Similarity 20.4%; Pred. No. 0.026;
Matches 176; Conservative 109; Mismatches 269; Indels 309; Gaps 45;
QY 11 QENKDNRVSYVDSQSQSKSENL--TPDQVSOK-----EGIOAEQIVIKITQCGVTS 63
Db 264 QEKKD-----LTAELSLNLFIDPTTQKGNONVEVKLGGETVSKIFNIQILGV 313
QY 64 GDHYHY-YNGKV---PYDALFSEELMKDPNYQLKDADIVNEVKGYILKVDG---KYV 115
Db 314 RDNKGVYTAGRIDTLNKVDGKFSHFAYMKPNNSQSLSVTVTGQTKGNKPGVNNPVKY 373
QY 116 VYLKDAHAADNVRTK-DEINRQKQEHVKDNEKYNVNAVARSQGRYTTNDGYVFNPA 174
Db 374 KHIGSDDLAESVAKLDDVS--KFEDVTDNNSLD-----FDINGGYSLENNL- 419
QY 175 EDTGNAYIVPHGGHYHYPKSDLSASELAARAAKHAHLAKGNMQPSQLSYSPSPSLP 234
Db 420 -DQSKNVIVIKEGY-----DSNASML-BFQTHLF-----YNYIYTSNLTWKN 463
QY 235 TSHEKHEEDGYGFDANRIIAEDSEGVSHGDNHYPFKKDLTEEQTKAAQKH-----LE 289
Db 464 VAFYSNNAQDGDGDK---LKEP---IIHSTPIELEFKSE-----PPVEKHHLTG 510
QY 290 EVKTS-----INGLDSLSHEDQYPSNAKEMKOLDKKEEKIAGIMKQYGVKRESIV 342
Db 511 ESNDKPIDPEYHTAVEGAECHAE-----GTIE-----TEEDSIHV 546
QY 343 NKEKNAIYPHGHHDADPT--DEHKPVGICHSHSNYELFKPEEGVAKK----- 389
Db 547 DFEEST--HENSXKHADVVEEDTNPGGQVTTESNLVEFDEDTKGIIVTGAUSDHIT 603
QY 390 -EGKVVITGELTNVNL-----LKNSTFNQNET---LANGQRVSEFSP 431
Db 604 IEDTKYETTE--SNLIELVDLPEEHQAQGPTEETIENHHSGLGTENGHNGYVI 661
QY 432 PELEKKLGINMLKLIITPDCKVLEKSVGVFGVGNIANFELDQPYLPQOTEKYTIASK 491
Db 562 EELEENSHVDIKSEL-----GYEGGQNSGN-QSFEED-----TEE 695

```

Qy 492 DYPEVSDYDGTFTVPTSLAYKMASTIYFPFHAGDTYLRVNPQFAVPKGTDLRVKVFDEHF 551
      ||| |
Db 696 DKPYEQGGNT-----VDIDE-----DSVPQIHQSNN 722
      ||| |
Qy 552 GNAYLENNYKGEIKLPIKLNOGTTTAGNKIPVTMANAYLDNQSYTIVEYPILEKEN 611
      ||| |
Db 723 GNQSFEDTEKDK-----PKYEQG-----GNLIIDFDSVPHIHGFNKH---TEIIIEDT 769
      ||| |
Qy 612 QTDKPS-----LPOFK-RNKAQENSKLD-----EKVEEPEKTS 643
      ||| |
Db 770 NKDXPNYQGGHNSVDFEEDFLPQVSHNEGQQTIEEDTPPIVPTPTPEVSPSEETP 829
      ||| |
Qy 644 EKVEKEKLSNGNSTSLTEVEPT----VDPQOEKVAKEAESYGMKLENLVNMDGTI 698
      ||| |
Db 830 TPTPEVPSBP--ETPTPTPEVTEPCGKPIPPAKEPEPK----- 867
      ||| |
Qy 699 ELYLPSGEVIKKNMADFTGEAPQNGENKPSNGKSVGTGVE-NQ-----PTENKPA-- 749
      ||| |
Db 868 ----PS-----KPVEQCKVVTPIEINKEKVAVVPTPKAQSCK 901
      ||| |
Qy 750 DSLPEAPNEKPKPENSTDNQML 772
      ||| |
Db 902 SELPETGGEE-----STNNQML 918
      ||| |

RESULT 14
B98002
Iga-specific metalloendopeptidase (EC 3.4.24.13) [imported] - Streptococcus pneumoniae
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence,revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: B98002
E: Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.;
  e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
  J. Bacteriol. 183, 5709-5717, 2001
A: Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A: Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A: Reference number: A97872; MUID: 21429245; PMID: 11544234
A: Accession: B98002
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-1963 <KUR>
A: Cross-references: GB:AE007317; PIDN:AAK99846.1; PID:g15458662; GSPDB:GN00174
C: Genetics:
A: Gene: iga
C: Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase
C: Keywords: hydrolase; metalloproteinase

```

Query Match	4.4%	Score 194.5;	DB 2;	Length 1963;
Best Local Similarity	22.0%	Pred. No. 0.098;		
Matches 165;	Conservative	85;	Mismatches 220;	Indels 281;
				Gaps 44;

QY 186 GGHHYIPKSDLSELAAKAHLAGNMQPQLSY-----SSTP 225
 ||| : ||| : |||
 Db 47 GVHKYVADELSEEL-----KKQLVDIPTVYVYKLSNQ 92

226 SPSPLPNGTSHEKHE-EDGYCFDANRIAEDESGFVMSHGDNHYFFKKDLTEEQIKAA 284

```

DD 285 QKHLLEVKTSNG-----LDLSH-----EQDPSNAKEMKDLDKKIE-EKIAGIMKQYG 334
93 LAELF-NIGSNERUALVAGASLAALGILI-----FAVS-----KKVKRNKIV--- 133

```

Db 135 -LHLVLVAGMGNGLVSVHALENHLLNNTDYELTSSEKLPUPKEISGTYIGIKEGK 193

QY 335 VKRESIVNKEKNAI-----TYPH-GDH-----HHADPIDEHKPVGIG--- 371

Db 194 TTSDFEVSNQEKSAATPTKQOKVDYNTVNFVDHPSTVQAIQEQTPVSSIKPTFVQWEK 253

QY 372 -----HSHSNYELPKPEGVAKKGGNKVYT-----GEELTNV 404

```

254 PFSTELINPRKEEKQSSDSQEQLAEHKNLET-KKEEKISPKEXTGVTNLPQDEVLSGQL 312
Db
04 405 N-----LIKSTEN-----NONETIANGCFDVSTGSDDEETPVKICINM-LVYVIT-- 448

```

[illegible]

Db	313	NKPELLAREETIERKIDFOEEIOENPDLAEGTVRV-----KQEGKLGKKVEIRIFSVN	366
Qy	449	-----PGKVLEKVSCK--VFGE-----GVT-----GNTANFELDOPIYP	480
Db	367	KEEVREIVSTSTAPSPRIVERIGKTKQIVKEQPTGHEKHDVQSGAIVEPAI-OPELP	425
Qy	481	GOTKYTIASKDYPEVSDGTFTVPTSLAYKMASQITIFVFFHAGDNYLRVNP0----	536
Db	426	-----EAVVSDKGEFVOP-----TLPEAVV-----TDKGEI-EVQESPTTVVS	464
Qy	537	PKGTDALRVFDEFHGNAYLENNYKVGEEK--LPIPKL-NQGTTRTAGNKIPVTEMANAY	593
Db	465	DKGEPEQVAPLPYKGN-----IFQVAPETPVEKTRKQSGPEKT--EEVPV-----	507
Qy	594	LDNOSTVIVEVPILEKENQTDKSIILDPFARN---FAQENSKLDKVPKPKTSEKV----	646
Db	508	-----KPTETFPVNPNGTTEGTSI--QEAENPVQPAEESTTNSEKVSPTDSENTGEVS	560
Qy	647	-----EKEKLSTGTSNSTLEVEYPTDVPQEKVAKFAESYGMKLENVLEN	693
Db	561	SNPDSSTSVGESNKPSPHNSDKNSENSEKTVVEVP-VNP-----	597
Qy	694	MDGTIELYLPSGEVIKKNMADFTGEAPQNGENKPSGENKV---STGTVENOPTENKP--	748
Db	598	NEGATIVE-----GTSNQETKERQVPAEETQTNSGKIANTENGTGSYKNPDSKPPV	646
Qy	749	ADSIUEAPNEKPKVPKENS-----TDNGMINPE	775
Db	647	EESNQPEKNGTATPKPNSGNTTNGSGQTEPE	677

RESULT 15

H71621
serine/threonine-specific protein kinase (EC 2.7.1.1.) PFB0150C - malaria parasite (P1
C:Species: Plasmodium falciparum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: H71621
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: H71621
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2485 <GAR>
A:Cross-references: GB:AE001376; GB:AE001362; NID:g3845108; PIDN:AAC71820.1; PID:g384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0150C
C:Superfamily: malaria parasite serine/threonine-specific protein kinase PFB0150C; pr
C:Keywords: phosphotransferase
P:2087-2352/Domain: protein kinase homology <KIN>

Query Match 4.4%; Score 194.5; DB 1; Length 2485;
 Best Local Similarity 20.0%; Pred. No. 0.14;
 Matches 175; Conservative 120; Mismatches 316; Indels 265;

QY 5 LNCERSQENKDN-NRVSIVDGSQSOKSENTPDQVSOKEGI-----QAEQIVIKIT 55

d0 796 ITPYESDKENNSNLIKSMDFENVVRKSNLYNNLSNRDSTYDMHKNYSFEVINIORT 855

QY 56 DQGYVTSHG--DHYHYNGKVPYDALFSEELMKDPNPQLKADIVNEVKGYYIKVDGK 113

DD 850 NKIYELSNARIKINYKLYSDMEIFVSLKERYIDNISNMERVYTKNEINERISKMDDI 911

QY 114 YYVYLKDAHADNVRTKDEINRQKQEHVKD-----NEKVNSNVAVARSQGRYTTNDGYVF 168

Db 916 LYPCKNKSINMSCPVIENNISRENEKSNSSVILNKKNN--MFCVGRLHCHMGK- 972

QY 169 NPADIIETGN----AYIVPHGGHYVIPKSDLSASELAARAAHLAGKNMQPSQ---- 218

[illegible]

GenCore version, 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:50:17 ; Search time 9.85337 Seconds
(without alignments)
3535.857 Million cell updates/sec

Title: US-09-471-255-16

Perfect score: 4396

Sequence: 1 CAVALNQHRSOENKNNRVS.....GTIELRLPSCVKKNLLIS 840

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186.5	4.2	1849	1 IGA4_HAEIN	P45386 haemophilus
2	180	4.1	1435	1 EBAL_PLAFC	P19214 plasmodium
3	175.5	4.0	1164	1 BAG_STRAG	P27951 streptococc
4	174	4.0	1139	1 HNW1_MYGE	Q49413 mycoplasma
5	173	3.9	1183	1 CNA_STAAU	Q53654 staphylococ
6	172	3.9	710	1 LT78_ARATH	Q06738 arabidopsis
7	171	3.9	1018	1 FNBA_STAAU	P14738 staphylococ
8	168	3.8	1702	1 IGA2_HAEIN	P45384 haemophilus
9	165.5	3.8	1637	1 MRSP_STAAU	P80544 staphylococ
10	165	3.8	1541	1 IGA1_HAEIN	P42782 haemophilus
11	164	3.7	1545	1 IGA3_HAEIN	P45385 haemophilus
12	163.5	3.7	1694	1 IGA0_HAEIN	P44969 haemophilus
13	162.5	3.7	715	1 STC2_STAAU	P17855 staphylococ
14	159.5	3.6	2195	1 SC16_YEAST	P48415 saccharomyc
15	158.5	3.6	1701	1 MSPI_PLAFC	P08569 plasmodium
16	157.5	3.6	3924	1 ANK2_HUMAN	Q01484 homo sapien
17	157	3.6	1381	1 YBE7_YEAST	P34216 saccharomyc
18	156	3.5	1453	1 Y373_BOVIN	Q9tu23 bos taurus
19	155.5	3.5	1701	1 MSPI_PLAFC	P13819 plasmodium
20	155	3.5	1073	1 RESA_PLAFC	P13830 plasmodium
21	154	3.5	1750	1 Y832_METJA	Q58242 methanococ
22	154	3.5	1790	1 USO1_YEAST	P25386 saccharomyc
23	153.5	3.5	1577	1 HLYA_PROMI	P16466 proteus mir
24	153.5	3.5	1726	1 MSPI_PLAFC	P04934 plasmodium
25	153.5	3.5	1726	1 MSPI_PLAFC	P50495 plasmodium
26	153.5	3.5	2748	1 NUM1_YEAST	Q00402 saccharomyc
27	153	3.5	1744	1 TANA_XENLA	P01550 xenopus lae
28	152.5	3.5	1630	1 MSPI_PLAFC	P04932 plasmodium
29	152.5	3.5	1639	1 MSPI_PLAFC	P04933 plasmodium
30	152.5	3.5	1658	1 YM67_YEAST	Q03661 saccharomyc
31	152	3.5	747	1 Y030_UREPA	Q9ptb5 ureaplasma
32	149.5	3.4	857	1 NFM_CHICK	P16053 gallus gall
33	148.5	3.4	1159	1 N124_SCHPO	Q09904 schizosacch

34 148.5 3.4 1287 1 VAC2_HELPY Q48245 helicobacte
35 148.5 3.4 1466 1 SPA2_YEAST P23201 saccharomyc
36 147.5 3.4 1167 1 SCAL_STRPY P15926 streptococ
37 147.5 3.4 1233 1 YE16_YEAST P43597 saccharomyc
38 147.5 3.4 1433 1 REST_CHICK O42184 gallus gall
39 147 3.3 818 1 PTK2_YEAST P47116 saccharomyc
40 146.5 3.3 3135 1 S230_PLAFC Q08372 plasmodium
41 146 3.3 1085 1 IFH1_YEAST P39520 saccharomyc
42 146 3.3 1240 1 YN1_YEAST P53935 saccharomyc
43 146 3.3 1332 1 SPT7_YEAST P35177 saccharomyc
44 146 3.3 1636 1 BUD3_YEAST P25558 saccharomyc
45 145.5 3.3 774 1 GLYE_PLAFC P02895 plasmodium

ALIGNMENTS

RESULT 1
IGA4_HAEIN STANDARD; PRT; 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision: Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTI HK61;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC -!- PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
certain Pro-|-Xaa bonds in the hinge region. No small molecule
substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M87491; AAA24968.1;
DR MEROPS; S06.001;
DR InterPro: IPR000710; Iga_S6.
DR InterPro: IPR004899; Pectact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pectactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 299 299 PROBABLE.
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AFA CRC64;

```
Query Match          4.2%  Score 186.5;  DB 1;  Length 1849;
Best Local Similarity 18.0%  Pred. No. 0.12;
Matches 166;  Conservative 128;  Mismatches 344;  Indels 283;  Gaps 35;

QY 12 ENKDNRRVSYVDGSO--SSCKENLTPDOVSOKEGIOAEQIVIKITDQGYVSHGDHYY 70
Db 471 KGNKGLLKGVDGTVILKOKADANKKQVAFQVGVISGRSTLVNDKQVDPNSYIFGR 530
QY 71 NCKVPYDA-----LFESEELMKDPN-----YOLKADADI 98
Db 531 GGRDLNGSLTFDHRNIDDGARVVNHNMTNITITGESLITNPWTITSYILEAQDD 590
QY 99 VNEV-----KGGYIIKVDGKYVYLKDAHADNV-----RTKDEINRQ 136
Db 591 DPLRIRSIYRQLYFNQDNRSYITLKRGASTRSELPONGSESNNWLMYGMTSDEAKRN 650
QY 137 KOEHVKNKNSVAVARSQRYTTNDGYVFNPAIIEDTGNAYIVPHGGHYHYIPKSD 196
Db 651 VNNHI--NNRMNG-----FNGYGEETKATQNGKLVNTNG-----KSD 689
QY 197 LSASELAATAKAHLACKNMOPSSOLSYSSSTPSPSLPINPQTSHEKHEEDCYGFDANILAE 256
Db 690 QNRFLITGTLNGDLNVEKGTFLSGRPTPHARDIAGISSTK--KDPHFTENNEVVVED 747
QY 257 E-----SGFVMSHGDHNNHYFPKDKLTBEQIKAAQKH 287
Db 748 DWINRFKATTNVNTGNSLYSGRNVANTSNITASNAAQVHIGYK---TGDTVCVRSDY 804
QY 288 LEEVKTSHNGLDLSLSHEODYPSNAKEMKDLKKI-----EKIAGIMKQYGVKRESIVW 342
Db 805 TGYV--TCHNSLSEKALNSFNENLNGVNLTENASFTLGRKANLFGTICQISGTSQ---V 859
QY 343 NREKNALIIPHGDHHAIDIDEKPKVGI--GHSHSNYELFKPEEGVAKKGNKVTGEEL 400
Db 860 NLKENS-----HWHLTGNSVYNOLNLTNGHILN-----AONDANKVITYNTL 902
QY 401 TNNVLLKKNSTF-----NNQNFPLANGQRVSPSPPELEKRLGIMLVLKLTIPDGK 452
Db 903 T-VNSLGSNGSFYVYVDFTNKSNKVVN--KSATGNFTLOVADKTPGPNHNLIFDAS 959
QY 453 VLEKVGKVGEGVGNIANFELDQYLPQTEKYTIASKD-----YPEYSYDGTFTVPT 506
Db 960 NATRNLEV-----TLANGSVDR-----GAWKYLRNNGRYDLYNPEV-----998
QY 507 SLAYKKAQOTIFYPHFAGDTYLRVNP--QFAPVPGTDALVRVDFEFGHGNAYLENNYKV 563
Db 999 -----EKRNQTV-----DTNITITPNDIQADAPSAQ-----SNNEETA 1031
QY 564 EIKLPKLNQGTTRTAGNKIPVTFMANAYLONQSTYIVVEVILKENQTDKPSILPQFK 523
Db 1032 RVETPPVPPAPATESAIAEQEPT-----RPAETAQPA-----1064
QY 624 RNKAQENSKLDEKVEEPKTSKYEKEKLSSETGNSSTNSLTLEVPVTPDQVQKVAFAESY 683
Db 1065 -----METNFTANSTETAPKSDTATOTENPNSVPS-----ETTEKVAENP 1106
QY 684 GMKLENVLFNMDGTIELYLPSSGVIFKKNMADF-----TGEAPQGGENKPSNGKVSIGT 738
Db 1107 POENETVAKNEQATEPTPQNGEVAKEDQPTVEANTQTNEATQSECKTEETQ-----TAE 1161
QY 739 VENQPT-----ENKPADSLPEAPNEKPV-----KPNSTDNGLMNPENGVSDPM-----783
Db 1162 TKSEPTESVTVSNQPEKTVSQSTEDRVVVVEKEKAKVETEETQKAPQVTSPEPPQAE 1221
QY 784 -----LDSALEEAPAVDPVQ 798
Db 1222 APEVPTDITNAEEAQAALQQTQ 1242

RESULT 2
EBAL_PLAFC
ID EBAL_PLAFC
AC P19214;

PRT: 1435 AA.
```

```
01-NOV-1990 (Rel. 16, Created)
01-FEB-1996 (Rel. 33, Last sequence update)
01-FEB-1996 (Rel. 33, Last annotation update)
DE Erythrocyte-binding antigen EBA-175.
OS Plasmodium falciparum (isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID:5835;
RN SEQUENCE FROM N.A.
RX MEDLINE:90377299; PubMed:2204835;
RA Sim B.K.L.;
RT "Sequence conservation of a functional domain of erythrocyte binding
antigen 175 in Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 41:293-296(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52524; CAA36756.1;
CC PIR: S11561; S11561.
CC DOMAIN 159 1104 ESSENTIAL FOR BINDING TO
FT ERYTHROCYTES.
FT VARIANT 1031 1031 E -> V (IN STRAINS FCR-3 AND ITG).
FT SEQUENCE 1435 AA; 167389 MW; 32A4309021B1C3D6 CRC64;
SQ
Query Match 4.1%  Score 180;  DB 1;  Length 1435;
Best Local Similarity 19.3%  Pred. No. 0.18;
Matches 113;  Conservative 95;  Mismatches 197;  Indels 182;  Gaps 26;

QY 11 QENKDNRRVSYVDGSOSSQSKSENLTDP-----QVSKQEGIOAEQIVI--KITDQGYVISH 63
Db 789 QKERDDSLSKISVSPNSRPETDAKOTSNLLKLGVDISMKPAKAVIGSSPNDINIVTEQ 848
QY 64 GDHYHYNGKVPYDALFSEELLMDPNYQLKADIVNEV-----102
Db 849 GDNISGVNSKPLSDVDPDKKELEDQNSDESEETVNVNHSKPSINNGDDSGSATVSE 908
QY 103 ----KGGYIIKVD--GKYVYVLKDAHADNVRTKDEINROKQEHVKNKNSVAVARS 156
Db 909 SSSNTGLSIDDDRRNGDTFVRTQDTANTDEVIRKENACKDEKGADEERHSTSESLS 968
QY 157 QGRVTTND--GYVFNPAIIEDTGNAYIVPHGGHYHYIPKSDLS-----ASELAATA 207
Db 969 EEKMLTDNEGNSLNHEEVKEHTNSDNVQSGGIVNNVNEKELKDTLENPSSSLDEGKA 1028
QY 208 H--LACKNMQPSQSYSTPSPSLPINPGTSH-----1083
Db 1029 HEELSEPNLSDQ--DMSNTPGPL-----DNTSEETTERISNNKYNREDERITLKKEYED 1083
QY 239 -----KHEEDGVGFDAIRIIA--EDESQV-----MSHGDRNHFF--KK 274
Db 1084 IVLKSHMNPESDDGELYDENSDDLSTVNDESEDAEAKMKGNDTSEMNSHSHIESDOCKN 1143
QY 275 DL-----TEQIKAAQK-----HLEEVKTSUNGILDSLSHEDDY 308
Db 1144 DMKTVGDLGTHVQNEISVPVPTGETIDEXKRESKESKIKHAEERLSHTDTHKI--NEPED 1201
QY 309 PSNAKEMKDLCKTEEKIAGIMKQYGVKRESIVVNNKKNATIIYPHGDH-----HADPIDE 364
Db 1202 NSNTLHLDIRNEENER-----HLTNQINISQED--LQKHGFTNNLHGDGVSE 1251
QY 365 HKPVGIGHSHSNYELFKPEEGVAKKGNKVTGEBLTNVNLLKNSTFNQNFNLANGQK 424
Db 1252 RSQ--INHSH-----HGNRQDRGGNSGVNLNMR-----NNNFN--NIPS 1288
QY 425 RVYSFSPPELEKLGKGINM-----LVKLTIPDGKVLKYSVK 460
```


[illegible]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Pritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
RA "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 721-847 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
RA "A survey of the Mycoplasma genitalium genome by using random
RT sequencing."
RL J. Bacteriol. 175:7918-7930(1993).
CC - FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).
CC - SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
CC MEMBRANE (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U39712; AAC71534.1;
DR EMBL: U02261; AAD12527.1;
DR TIGR: MG312;
KW Cytadherence; Structural protein; Complete proteome.
SQ SEQUENCE 1139 AA; 130531 MW; 0011D3288C3DD856 CRC64;

Query Match 4.0%; Score 174; DB 1; Length 1139;
Best Local Similarity 19.9%; Pred.No. 0.26;
Matches 169; Conservative 117; Mismatches 305; Indels 260; Gaps 45;

Qy 70 YNGRVYDALFSELLMKDPNYOLK-DADIVNEVKGYYII---KVDGKYVYVYKDAHAD 125
Db 27 YDGK-----STVEDPNKIVAYDAD-----GNGYIAFNKRTGVYDYPYGDTEY-D 71

Qy 126 NVRTKDE-----INRQKEHVKNKNSVAVARSOGRYTTNDGYVFNADIETGNA 190
Db 72 ISQLFDENGPNFVDFEKOEE-NVLYKTVGN-----PDGYSDENGEWNWS----- 115
Qy 181 YIVPHGGHYHYIPKSDLSASALAAKHAHLAKGNMQPSLSSTPSLPIPGTSHEKH 240
Db 116 -----GVF-----END-----QWISTKESQPTDENYGF--DSDLF--PEVKOPES 151
Qy 241 EEDGYGFANRI-----IADESEF-----VMSHGDHNYFFKKDLTEE 279
Db 152 VEDNYGFD-NDLPPPEVKOPESVEDNYGFDNDLPEVKOPESVVDQPSDDYFAKQP-TDE 209
Qy 280 QIKAAQKHLBVKVSHNGLOSLS-----SHEQDYPSSNAKEMKDLQKIE 323
Db 210 NYGFDNDLPPEVKOPESVVDQPSDDHFAKQPESTTDSYSDSLDLPOTLDQPSLDHVV 268
Qy 324 EKIAGIMKQGVKRESIVVKNKNAIYIPGDHHADPDEHPVGTGSHSNVLFKPE 393
Db 269 -----QYNF-----DHH-----EELKPV--AEQNNYQV----- 290
Qy 384 EGVAKKEGNKYVTGTELTNNVNLKNSTNNONFTLANGQKRVSFSPPELEKRLGINML 443
Db 291 -GFDOVQAN-LDNEETOPTAEKVVITDFESKQAVVD-----SYQLPIDTDQDDQTFS 343

Qy 444 VKLIT-PDGKLVKSGKVGFGVGNIANFELDQPLPGOTFKYTIASKDYPEYSDGTF 502
Db 344 SSFETOPTVEFQDVNSEV-----NQQKPEITKEPVLESFKNQD 384
Qy 503 TVPTSLAYKMASQIFVYFFHAGDITYLRVNPQFAVPKGTDALRVF-----DEFH 551
Db 385 VVETS---DLNSESNLISENNKDA-----TNDLSLNSSEFIQLNSSETASDDVH 430
Qy 552 GNAVLE--NNYKVGKILPIPKNOGTRTAGNKIPVTFMA---NAYLDNQSYI--VE 603
Db 431 YESKSEPIHDYKFGS-----DLSQSNNSNLSSEPKFNSSETAPDAHFESQSPVQVQ 484
Qy 604 VPILEKEN---QTDKPSILPQFKRKAQENSKLDEK-----VEPKTSEKVEKE 649
Db 485 YDIYQNEELKPTLDQPSDDYFAKOPTDENYGFNDLPPPEVKOPESVVDQPSDDHEAKQ 544
Qy 650 KLSGTGNSNSTLEEVEVTPD-----VQ-----EKVAKFASYSGMKLENVLFNM 694
Db 545 PESTTDSYSFSDLPQ-PTLDQPSLDHVVQNFHHEELKPVAEQNNYQVGFQVQANL 603
Qy 695 DGTIELYLPSSGEVIKKNMADFTGEAPQNGENKPSKENGKVTGTVENQPTENKPADSLPE 754
Db 604 DNNEEIQ-PTAE--KEVTTDFESKQAQ-----VVDSYQLPIDT--DQDQDTFESSFET 652
Qy 755 APNEKVPKPESTNDGMLNPEGNVGSDDPMLDSALEAPVDPVQEKLEKTASYGLGLDS 814
Db 653 QPTVQFQDVANSEVNDQFKPE--ITKEPVLESFKNQDVE-----TSNTNNLQK 701
Qy 815 VIFNMDGTIEL 825
Db 702 FDIQSDNKITI 712

RESULT 5
CNA_STAAU STANDARD; PRT: 1183 AA.
ID CNA_STAAU STANDARD; PRT: 1183 AA.
AC Q53654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.,
RT "Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin."
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.,
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoeoek M.,
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus."
RL Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carlson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., DeLucas L.J., Hoeoek M.,

[illegible]


```

CC      or send an email to license@isb-sib.ch).
CC      EMBL: M87489; AAA24566.1; -.
DR      MEROPS: S06.001; -.
DR      InterPro: IPR000710; IGA_S6.
DR      InterPro: IPR004899; Pertact_sup.
DR      Pfam: PF02395; IGA1; 1.
DR      Pfam: PF03212; Pertactin; 2.
DR      PRINTS: PR00921; IGASERPTASE.
KW      Hydrolase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
FT      SIGNAL.
FT      CHAIN.
FT      PROPEP.
FT      ACT_SITE.
FT      DOMAIN.
FT      REPEAT.
FT      REPEAT.
SQ      SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

Query Match
Best Local Similarity 18.8%; Score 168; DB 1; Length 1702;
Matches 175; Conservative 123; Mismatches 347; Indels 286; Gaps 42;

QY      12 ENKNNRVSYVDGSG--SQSKSNLTPDQVSQKEGQIAEQYIKITDQYGVTSVSHGDHYHY 70
DB      467 ENKSLKVG--DGTILKQADANNKAFQSGVIGSRSTVVLNDDKQVDPNSYIFGFR 524
QY      71 NGKVPYD-----ALFSEELMKDPN-----YOLKQDADI 98
DB      525 GERLDANGNLTPEHNRIDDCARLVNHTSKTSVITVTGESLITDPTIIPYNIDAPDE 584
QY      99 VN-----EYKGG---YIKVQDGYVYVYLKDAHADNV-----RKDEINR 135
DB      585 DNPYAFRIKDGQGLVNLNENTYVYALRKGASTRSELPKNSGESNENLYMGKTSDEAKR 644
QY      136 QKQEHVKDKEKNSNVAVARSQRTTNDGVYFNPADIETDGNAYIVPHGGHYHIFKS 195
DB      645 NVNHHI--NERNWG-----FNGYGEIEG-----KNGNUNVTFKG-----KS 681
QY      196 DLSASELAAKAHAKGNMPSQSYSTPSPSLPINSPTGSHKHEEDGYGFDANRIAE 255
DB      682 EQNRFLLTGTTNLGDLKVEKGTFLSGRPTPHARDIAGISSTK--KQHFENNEVVVE 739
QY      256 DESGFVMSHGDHNYFFKK--DLTEEQTKAAQKHLVEVKTSNGLDLSLSS----- 304
DB      740 D-----DWINRFKATNIVTNNTATLYSGRNVANITSNITASDNKAVHIGYKAGDT 790
QY      305 ---EODYPSNAKEMKDLKKIEKIAGIMKOYGVKRESIVNKEKNA----- 348
DB      791 VCVRSYDYOYVCTTD--KLSDK---ALNSFNATVSGNVNLGNFVLCGKANLFGTI 844
QY      349 -----IIPGCDHH--ADPIDEHKVPGLGHSNVELFPEEGVAKKGNKYVTGEE 399
DB      845 SGTGNSQVRLTENSHTLTCGDSNVQNLNDKQGHILN-----KQNDANKVTTTNT 894
QY      400 LTNVNLKLNSTFNNQNTLANQKRVSEFPPELEKKLGNMLYKLITPDGKVLKESVG 459
DB      895 LT--VNSLS-----GNG-----SPYLTDLNKGDKRVVTK-SATGNFTLQVAD 936
QY      460 KVFGEVGNIANFELDQPLPGQTFKYTTASKDYDEPVSVDGFTVPTSLAYKMASQTIYF 519
DB      937 KT-GEPTKN-----EL-----TLFDASNATRNLSLVGNTVDLGAWKKYLRNVNGRY 984
QY      520 PFHAGTYLRVNPQ-----FAPYKGTDLVRVDFEFGHNAVLENNKYKVEIKL 567
DB      985 -----DLY--NPEVEKRNQTVDTNITTPNNIQADV-----PSVPSNNEELARVET 1028
QY      568 PIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVVEVPILEKENOTDRPSILPQYKRNKA 627
DB      1029 PVPP-----PAPAT-----PSETTETV-----AENSKQESKTVKEKNEQDAT 1064
QY      528 QENSKLDEKVEEPKTSK--VEKEKLSETGNSTSLTEEVFTVPVQ-EKVAKFAESYVG 684

```

Db 1065 ETTAQNGEAAEEAKPSVKANTQTNEVAQSGSEETQTTEIRKETAKEKEKAKVEKEK 1124
 QY 685 MKLENVLNMDGTIELYLPSEGEVIKKKNMADFTGEAPQNGENKPK-----SENGKV 734
 Db 1125 AKVEK-----DEIEAFQMASETSPKQAKPAPKEVSDITKTV 1160
 QY 735 STGTIVENQP-----TENKPADSL--PEAPNEKPKVPK---ENSTNDGMLNPE 775
 Db 1161 EETQVQAOPQTQSTTVAAAEATSPNSKPAEETQPSKTAEPVTPVYSKNOTENTTDQPT 1220
 QY 776 GNVGSPMLDSALEEAPV---DPVQEKLE 802
 Db 1221 EREKTAKVETEKTOEPPQVVASQASPKQEOSE 1251

RESULT 9
 MRSP_STAAU STANDARD; PRT; 1637 AA.
 AC P80544; Q92F62;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Methicillin-resistant surface protein precursor.
 GN PLS.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate 1061;
 RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;
 RT "pls, a large repeat-rich surface protein of methicillin resistant
 RT Staphylococcus aureus";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 624-628; 676-682; 938-948; 1156-1168; 1176-1185;
 RP 1199-1205 AND 1217-1224.
 RC STRAIN=Isolate 1061;
 EX MEDLINE=36270743; PubMed=8665912;
 RA Hilden P., Savolainen K., Tynnelae J., Vuento M., Kuusela P.;
 RT "Purification and characterization of a plasmin-sensitive surface
 RT protein of Staphylococcus aureus";
 RL Eur. J. Biochem. 236:904-910(1996).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF115379; AAD09131.1; -
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR TIGRFAMs: TIGR01167; LPXTG_anchor; 1.
 DR TIGRFAMs: TIGR01168; YSIRK_signal; 1.
 DR PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
 DR Antibiotic resistance; Glycoprotein; Cell wall; Peptidoglycan-anchor;
 KW Repeat; Signal. 1 48
 FT SIGNAL
 FT CHAIN 49 1601 METHICILLIN-RESISTANT SURFACE PROTEIN.
 FT PROPEP 1602 1637 REMOVED BY SORTASE (POTENTIAL).
 FT DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-(SAG).
 FT SITE 1598 1602 LPXTG SORTING SIGNAL (POTENTIAL).
 FT MOD_RES 1601 1601 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 SQ SEQUENCE 1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;

Query Match 3.8%; Score 165.5; DB 1; Length 1637;
 Best Local Similarity 20.2%; Pred. No. 1.1;

MATCHES 168; Conservative 88; Mismatches 327; Indels 247; Gaps 39;

QY 120 DAADNVRKDEINRQKHVKD-----NEKVNSNVAVAR-----SOGRYTNDGY 166
: : : : :
Db 323 EKABIEKVLPKD- ISNLSNEIKKIALSEVLKETANKENAQPRATERSVSSNARTNY 381
: : : : :
QY 167 VFN-----PADLIEDGNAYIVPHGHHYHYPKSDLSASELAHAHLAKKNMPS----- 217
: : : : :
Db 382 SATALRAAAGDTVTKGTGNTFAGDLIHKTKEFFPNECTLTA-----FTNFNFNTGTK 437
: : : : :
QY 218 -----QLSYSTSPSPSLPINPTGSHEKHEEDGVGFDAARRIAED-----ESGEVMSHDH 267
: : : : :
Db 438 GALEYNKDIDENKDFITIPVANNNOGNTTGADGWGMEQTQNGDQFLNQGLIRDKMA 497
: : : : :
QY 268 NHFFFKDLTEBQKAQKHLEBKVTSHNLDSLSHEDYPSNAKEWKDLKTEEKIA 327
: : : : :
Db 498 NASGFIKIDTAYNNVNGVKDKLDADKT--NNLSQIGAAKYGYTFPVKGAD----- 545
: : : : :
QY 328 GLMKQYGVKRRESIWNKEK-NAIIVPHGDHHDAPIDEHKPVGIHSHS-----NY 377
: : : : :
Db 546 GVYNONG---ONALNFKDPVNKIIVADNTTNHLD-----GQHGGRLNDVVUNY 592
: : : : :
QY 378 ELFKPEEGVAKKGNKYVEELUTNVNLLKNSTFNNOQNTLIANGCKRVSPS-----FPPEL 434
: : : : :
Db 593 D-----AATSTITATYAGTKWAKTTDGLGDKSKVNFILTSSTMQNRYSGIMRTNL 645
: : : : :
QY 435 EKKLGINMLVKLTPDGKLVLEKYSKVFGEGVGNIANFELOQPYLPQGTYFYTASKDYP 494
: : : : :
Db 646 EG-----VTITPPADLIDDV-EVTQPIPHKTIREFDTLPFS-----P 685
: : : : :
QY 495 EV-----SYDG--TFTVPTSILAVKMASQTIYPFHAGDYLRVNPQFAVPKG-TDALVRVF 547
: : : : :
Db 686 DVIVQKGDEKTTTTPT-----KVDPD-----GDVVERGEPTTEVTKNPVDIVHT 734
: : : : :
QY 548 -----DEPHGNAYLENNYKVGEEKPLPKLNQGGTRTAGNKIPVTEMANAYLDNQS 598
: : : : :
Db 735 PEVPQGHKDEFDPN-----LPI-----DGTEEVPG-----KPGIKNPE 768
: : : : :
QY 599 TYIVEVPILEK-----ENQTKPSILPOFKRNKAQENSCLDERVEEPTSEKVEKE 649
: : : : :
Db 769 TGEVTPVPDDVTKHGPKAGEPEVTEKEETPEKKEFPNPLKPGEKV-----TQEGQTGE 824
: : : : :
QY 650 KLSETGNSNSTLEE-----PTVDPVQKVAKFAESVGMKLE-----NVLFNMDCGI 698
: : : : :
Db 825 KTTTPTTNPULTEKVGGEPTTEVTKFEVDEITQFGEEVPOQGHKDEFPNLPIDGT 884
: : : : :
QY 699 ELXLP-----SGEVKKMAFDFTGAPOCN-----GEN 726
: : : : :
Db 885 E--VPCKPGIKNETGEVTPVDDVTKHGPKAGEPEVTEKEETPEKKEFPNPLKPGE 942
: : : : :
QY 727 KPSENKSVTGIV-----ENQPT-----ENKPADSLPEANEK----- 759
: : : : :
Db 943 KVTOECGTGKTTTTPTTINPLTGEKVGGEPTTEVTEKEPEVDEITQFGEEVPOQGHKDEF 1002
: : : : :
QY 760 -PVKPENSTDN-----GMLNPE-GNVGSDPMJLDSALEEAPADVQEKLE 802
: : : : :
Db 1003 DPNLPIDGTEEVPGKPGIKNETGETVTPVDDVTKHGPKAGEPEVTEKE 1052
: : : : :

RESULT 10
IGAL_HAEIN
ID IGAL_HAEIN STANDARD; PRT; 1541 AA.
AC P42782;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGal protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OX NCBI_TaxId=727;
[1]

SEQUENCE FROM N.A.
STRAIN=HK368 / Serotype B;
MEDLINE=89379374; PubMed=2506130;
Paulsen K., Brandt J., Hjorth J.P., Thøgersen H.C., Kilian M.;
Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
of Haemophilus influenzae serotype b.";
Infect. Immun. 57:3097-3105(1989).
[2]
MUTAGENESIS OF SER-288.
STRAIN=HK368 / Serotype B;
MEDLINE=92234949; PubMed=1373717;
Paulsen K., Reinholdt J., Kilian M.;
A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
J. Bacteriol. 174:2913-2921(1992).
CC -! FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -! CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
certain Pro-I-Xaa bonds in the hinge region. No small molecule
substrates are known.
CC -! SUBCELLULAR LOCATION: Secreted.
CC -! DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROLYSIS (BY
SIMILARITY). BELONGS TO PEPTIDASE FAMILY S6.
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X64357; AAA45708.1; .
EMBL; M87492; AAA24969.1; .
DR MEROPS; S06.001; .
DR InterPro; IPR000710; Iga_S6
DR InterPro; TPR004899; Pertactin_sup.
DR Pfam; PF02395; IGAL; 1.
DR Pfam; PF03212; Pertactin; 2.
DR PRINTS; PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1008
FT PROPEP 1009 1541
FT ACT_SITE 288 288
FT MUTAGEN S->T; LOSS OF ACTIVITY.
SQ SEQUENCE 1541 AA; 169370 MW; CE7257CB3196C600 CRC64;
Query Match 3.8%; Score 165; DB 1; Length 1541;
Best Local Similarity 19.3%; Pred No. 1.1; Matches 172; Conservative 120; Mismatches 120; Indels 268; Gaps 44;
QY 12 ENKDNRRSVYVPGSSQSKSENLTPOVSQKEIOAEQIVKIQDQGVTVTHGHYHYHN 71
Db 462 DNKGLKVG--DCTVILKQQTNGSGOHAFASVGVSGRSTVLVNDKQVDPSNIYVGRG 519
QY 72 GKVPYDA-----LFSEELLMDPN-----YQLKDAIV 99
Db 520 GRULDNGNSLTDFHNRINDDGARLVNHNNMTNASNTITGESLTDPTTTPYNDAPDED 579
QY 100 N-----EVKGG---YIIKVDGKYVYVYLKDAHADNV-----RTKDEINRQ 136
Db 580 NPVAFRIKGGQGLVNLNLEYIYVALRKGASTRSELPSKNGSNENLMYMKTSDEAKRN 639
QY 137 KOEHYKDNKENVN-VNAVARSQGRITNDGVFNPAIDIETGNAYIVPHGHHYHYTPKS 195
Db 640 VMNHI--NERNMGNGFYEGEECKNGNMLNTYFKG----KSEQRFLLTGCTNLN----G 690
QY 196 DLSSASELAALKAHLAGKNMQPSOLSVSTPSPSLPINPTGSHKHEEDGVGFDAARRIAE 255

DB 968 LGAWKYKLVNGRY-----DLY---NPEVEKRNQTVDTTITTTNNIQAQVSPVSPNNE 1019
QY 542 ALVRVDFEHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLNDQSTYI 601
DB 1020 ETARV-DE-----APVPPAPATPSETTETVAEN-----SKQESKT 1054
QY 602 VEVPTLEKENOTD-----KPSILPQFKNK-AQENSKLDE---KVPEPK 641
DB 1055 VB-----KNQDATTQAQREVAKEAKSVNKANTQINEVAQSGSETKETQTETKETA 1108
QY 642 TSEKVEKELSTGNTSNTLEEV-----TVDPQEKVAKFAESYGMKLVNLFNMDGT 697
DB 1109 TVEKEEKAHV-ETEXT-----QEVPKVTSQVSPKQEQ-----SETVQPAEPAREN-DPT 1156
QY 698 IELYLPSGEVILKKNMADFTGAPOGNGENKPSNGKYSTG--TVENOPTENKPADSLPEA 755
DB 1157 VNIKEPQSQNTIADTEQPAKETSSNEQPVTESTVNTGNSVVEN-PENTTPATTOPTV 1215
QY 756 PNEKPKVDEN 765
DB 1216 NSESSNKKPN 1225
RESULT 12
ID IGAO_HAEIN STANDARD PRT: 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA OR IGA1 OR HI0950.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype D;
RA Wright A., Fishman Y., Tai F., Plaut A.G.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=9530630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.",
RL Science 269:496-512(1995).
CC -!- FUNCTIONING VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A
CC -!- PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X59800; NOT_ANNOTATED_CDS.
DR EMBL: U32779; AAC22651.1; .
DR MEROPS: S06.001; .
DR TIGR: HI0990; .
DR InterPro: IPR007010; IGA_S6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal;
KW Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT CONFLICT 253 254 EN -> GV (IN REF. 1).
FT CONFLICT 272 272 G -> A (IN REF. 1).
FT CONFLICT 464 464 G -> E (IN REF. 1).
FT CONFLICT 866 866 S -> T (IN REF. 1).
FT CONFLICT 1036 1036 A -> D (IN REF. 1).
FT CONFLICT 1074 1074 A -> G (IN REF. 1).
FT CONFLICT 1421 1421 A -> G (IN REF. 1).
FT CONFLICT 1545 1545 H -> T (IN REF. 1).
SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;
Query Match 3.7%; Score 163.5; DB 1; Length 1694;
Best Local Similarity 19.0%; Pred. No. 1.5;
Matches 180; Conservative 125; Mismatches 385; Indels 259; Gaps 42;
QY 12 ENKDNRVSYVDGSO--SSQKSENLTDPQVQSKQEGIAEQIVIKITQGYVTSHGDRHYHY 70
DB 467 ENKSLKVG--DGTVLKQQAQDANKVKAFSGVIGRSTVVLNDDQVDPNSIYFGR 524
QY 71 NGKVPYD-----ALFSEELMKDPN----YQLKDADI 98
DB 525 GGRLDANGNNLTFEHIRNIDGAPLVNHTSKTSTVTITGESLITDPTTIPNIDAPDE 584
QY 99 VN-----EVKGG---YLIKVDGKYVYVYLKDAHADNV-----RTKDEINR 135
DB 585 DNPYAFERRIKDGGQLYLNLENTYIALRKGASTRSELPKNSGESNENLWYMGTSDEAKR 644
QY 136 QKQSHVDNKEKSNVAVASQGRYTTNDGYVFNPAIDEDTGNAYIVPGHGHVHPKS 195
DB 645 NVNMHI--NERNMG-----FNGYFGEERG-----KNNGLNVTFKG-----KS 681
QY 196 DLSASELAATAKAGLAGNMOPQSLSYSTSPSLPINPQTSHEKHEEDGYGFDANRIAE 255
DB 682 EQNREFLLTGTLNGLDKVEKGTFLSGRTPHARDIAGISSTK--KDQHFAENNEVVVE 739
QY 256 DESGFVMSHGDHNYFFKK--DLTEEQIKAAQKHLEEVKTSHNGLDLSLSSH----- 304
DB 740 D-----DWINRNFKATNINVTNNATLYSGRNVANITSNTASDNKAVHYGKAGDT 790
QY 305 ---EQDYPSNAKEMKDLCKIEEKIAGIKQYGVKRESIVVNEKNA----- 348
DB 791 VCVRSDDTYGVTCITD---KLSDK---ALNSFNATNVSGNVNLSGNANFVLGRANLFGTI 844
QY 349 -----IIVPHGDHHH--ADPIDEHPKPVGIGHSHSNYELFKPEEGVAKKEGKVTGEE 399
DB 845 SGTONSQVRLETENSHWLTGDSNVNQNLNDKGHIHLN-----AQNDANKKVTYNT 894
QY 400 LTNVNLKNTNNQNFTLNGQKRVSFSPPELEKLGKGINMLVLKITPDGKVLKESVG 459
DB 895 LT--VNSLS-----GNG-----SFYLTDLNKKQGDVVVTK-SATGNFTLOVAD 936
QY 460 KVFEGGVGNIANFELDOPYLFGQTFKYTIASKQDYPEVSYDGTFTVPTSLAYKASQTFY 519
DB 937 KT-GEPTKN-----EL-----TLFDASNATRNRLNVLNVLGNTVDLGAWKYKLRNVNGRY 984

```
QY 520 PTHAGDTYLRVNPQ-----FAVPKGTDALRVFDEFHGNAYLENNYKVGEIKL 567
DB 985 -----DLY---NPEVEKRNQVDTTNIITPNQIADV-----PSVPSNNEETARVET 1028
QY 568 PIPKLNQGTTRAGNKIPVTFMANAYLNQSTYIIVEVPILEKENOTD-----KPSILPQ 621
DB 1029 PVPVPAPAPP-----SETTETVAENSKOESKTVKEKNEQDATETTATQAGVEAVEAEKPSVKAN 1084
QY 622 FKRNK-AQENSKLDE-KVEEPTSEKVEKELSGTNGSTNSLTLEEVPTV---DPVCEK 675
DB 1085 TOTNEVAQSGSTTEOTTEIKETAKVEKEKAK-----VEKDEIQEAPOMASETSPKOAK 1140
QY 676 VAKFAESYGMKLENNLFNMDSGTIELYPSGEVIKKNMADFTGEAPQNGENKPSN-----731
DB 1141 PAPKEVSTDTKVEETQVQAO-----POTQSTTVAAAEATSPNKPAAEETQPSSEKTNAE 1193
QY 732 -----GKYSTGTVENQPTENKPADSLPEAPNEKPKPENSTDNGLMNPENGVGSDPMLD 785
DB 1194 PVPVVSKNQNTENTDQPTER---EKIAKVFETKQETQEPQVVASQSPKQOSETVQPOAV 1250
QY 786 SALEEPAPV---DPVQEKLEKFTASYGLGLDSVIFNMDSGTIELRLPSGE 831
DB 1251 LSENNYPTVNAEEVQAOIQOTSA-----TVSTKQAPPE 1285

RESULT 13
ID STC2_STA00 STANDARD; PRT; 715 AA.
AC P17855;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Staphylocoagulase precursor.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=BB.
RX MEDLINE=88139269. PubMed=3481366;
RA Kaiba S., Miyata T., Yoshizawa Y., Kawabata S., Morita T.,
RA Igarashi H., Iwanaga S.;
RT "Nucleotide sequence of the staphylocoagulase gene: its unique COOH-
terminal 8 tandem repeats.";
RL J. Biochem. 102:1177-1186(1987).
CC -!- FUNCTION: STAPHYLOCOAGULASE IS AN EXTRACELLULAR PROTEIN WHICH
CC SPECIFICALLY FORMS A COMPLEX WITH HUMAN PROTHROMBIN. THIS COMPLEX
CC NAMED STAPHYLOTHROMBIN CAN CLOT FIBRINOGEN WITHOUT ANY PROTEOLYTIC
CC CLEAVAGE OF PROTHROMBIN.
CC -!- DOMAIN: THE C-TERMINAL TANDEM REPEATS ARE NOT REQUIRED FOR THE
CC PROCOAGULANT ACTIVITY.
CC -!- SIMILARITY: TO THE STAPHYLOCOAGULASE FROM STRAIN 213.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
Best Local Similarity 20.5%; Pred. No. 0.51;
Matches 154; Conservative 82; Mismatches 253; Indels 263; Gaps 33;

QY 124 ADVNRTKDEINRQKQEHVKNKVNNAVAVARQGRVTTNDGYVFNPAIDIEDTGNAYIV 183
DB 24 ADAIVTKD---YSKESKVNENSYGTILSDWYDKRLTSLSEQFINALDILET-----73
QY 184 PHGCHYHYIPKSDLSASELAAKHAHLAKGNQPSQSYSTSPSPSLPINPGTSHKHEED 243
DB 74 -----YHV-----GEKEYKD 83
QY 244 GYGFDAIRIIAEDSGFVMSHGDHNYFFKDLITERQIKAAQKHLEEVKTS-----294
DB 84 AKDKLMTRILGEDQ--YLLER-----KKQVYEYKLYQYKEENPTSKGLKLTED 133
QY 295 -----HNGL--DSLSSHEDQYPSNAK-----EMKDLKKIEEKIAGIM-----330
DB 134 QYTIEDITMREYNELTESLAKSAVKDFEKOVEKIENQHHDLKPTDEMEKATSRVDLAN 193
QY 331 KQYGV-----KRESIVWKEKNAILIYPHGDDHHAADIDEHKPVGIGHSHEWELF 380
DB 194 KAVSYVFAFVRDTHKTEALELRAKVDVLG-----DEKPHRISNERIEKMI 242
QY 381 KPREGVAKKEGNKVTYGTGEELTNVNLKNTENNQFTLANGOKRVSFSPPELEKKLGI 440
DB 243 KLESII--EDFIETGLNKPGNITSYSSKHHYKHS-----EGF 281
QY 441 NMLVKLITPDGKLVKSGVKGEGVGNIANFELDOPYLPQGTFKYTIASKDYEVSYDC 500
DB 282 EALVKETR-----EAVAN-----ADESWKTKTKVKGYESKTSPPVVKEN 321
QY 501 TFTVPTSLAYKMASOTIFYFFHAGDTYLRV-NQFAVPAKG--TDALRVVDFEFGNAY--555
DB 322 KVEDPQPKEDNQGEVKTATGAEETQVQAQVLKIQGTITGEIVK-----GPEIPT 375
QY 556 LENNYKVGEEKLPKLPKLNQGTTRTAGNKIPVTFMANAYL--DNQSTYIVEVPILE-KENOT 613
DB 376 MENKTLQGEI-----VQGPDPFTMQSGPSLSDNYTQTPTPNPILLEGSS 422
QY 614 DKPSILPQFRNKQAQ-----ENSKLDEKVEEPKTS-----KVEKEKLESG 655
DB 423 SKUEIKPQGTSTLKGIGQESSDIEVKPQATEATEASQYQPRQFNKTKPKYKIRDAGTC 482
QY 656 NSTSNS---TLEEVPTVDPVQERKAFKFAESYGMKLENNLFNMDSGTIELXLPSEVIKKN 712
DB 483 IREYNDGTGFGYEAPRPNKPSSETNA-----YVNTNQDGTVSY---GARPTQNK 528
QY 713 ADFTGE---APOQNG-----ENKPSE-----NGKVSSTGTVENQPTENKPAD 750
DB 529 ASETNAYNVTHANGQVSYGARTQKPKSETNAYNVTHANGQVSYGA---RPTYNKPS 585
QY 751 SLPEAPNEKPKPENSTDNGLMNPENGVGSDP 782
DB 586 T-----NAYNVITHG--NQQVSYGARP 605

RESULT 14
SC16_YEAST STANDARD; PRT; 2195 AA.
ID SC16_YEAST
AC P48415; Q02822;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Multidomain vesicle coat protein.
GN SC16 OR YPL085W OR LPL1W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96017704; PubMed=7593161;
RA Espenshade P., Gimeno R.E., Holzmacher E., Teung P., Kaiser C.A.;
```

"Yeast SEC16 gene encodes a multidomain vesicle coat protein that interacts with Sec23p".
J. Cell Biol. 131:311-324 (1995).
[2]
SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albertmann K., Allen E., Ansorge W., Araujo R., Aparicio A., Barrall B.G., Badcock K., Benes V., Bostein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M., Chung E., Churcher C.M., Coster F., Davis K., Davis R.W., Dietrich F.S., Delius H., DiPaolo T., Dubois E., Dueterhoeft A., Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A., Hall J., Hebling U., Heumann K., Hilbert H., Hillier L., Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K., Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J., Marathe R., Messing F., Mewes H.-W., Mirtipati S., Moestl D., Mueller-Auer S., Namath A., Newtich U., Oefner P., Pearson D., Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharie M., Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H., Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H., Walsh S.V., Wambutt R., Wang Y., Wedler H., Winnett E., Zhong W.W., Zollner A., Vo D.H., Hani J.,
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";
RT Nature 387:103-105(1997).
RL -!- FUNCTION: INVOLVED IN THE BUDDING OF TRANSPORT VESICLE FROM THE.
CC ENDOPLASMIC RETICULUM. THE C-TERMINAL INTERACTS WITH SEC23 AND
CC WITH THE CYTOSOLIC DOMAIN OF SED4. COULD THEREFORE BE A
CC CONSTITUENT OF COPII VESICLE COAT. N-TERMINAL OVEREXPRESSION
CC CAUSES A LETHAL SECRETION DEFECT.
CC -!- SUBCELLULAR LOCATION: ON THE ENDOPLASMIC RETICULUM AND ON VESICLES
CC WHICH BUD FROM IT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U23819; AAC49088.1; ;
CC EMBL: U41849; AAB68254.1; ;
CC SGD: S0006006; SEC16.
KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum.
FT DOMAIN 1997 2094
FT MUTAGEN 1059 1059 L->S: IN SEC16-4; TS ACCUMULATION OF ER
FT MEMBRANES.
FT MUTAGEN 1084 1084 L->P: IN SEC16-3; TS ACCUMULATION OF ER
FT MEMBRANES.
FT MUTAGEN 1089 1089 L->P: IN SEC16-2; TS ACCUMULATION OF ER
FT MEMBRANES.
FT MUTAGEN 1231 1231 W->R: IN SEC16-1; TS ACCUMULATION OF ER
FT MEMBRANES.
FT CONFLICT 522 522 MISSING (IN REF. 1).
FT CONFLICT 560 560 I -> F (IN REF. 1).
SQ SEQUENCE 2195 AA; 241694 MW; 757B7A7231BEBF0 CRC64;
Query Match 3.6%; Score 159.5; DB 1; Length 2195;
Best Local Similarity 21.5%; Pred. No. 3.3;
Matches 193; Conservative 97; Mismatches 347; Indels 261; Gaps 48;
QY 13 NKDNRVSVSDGSQSSQSENITPDQVSOKEGIAEQIVKITDQGYVTS-----HGDRYH 68
DB 1426 NKKNDKKVDFDFTGSSANSSTVLTQTTFPFAQ-----VTSQSYVDITALLNAH-- 1478
QY 69 YNKGVPYDALFSEBLLMKDPNQLKADIVNEVKGVIYKVDGYVYLKDA----- 121
DB 1479 ----NVP-----SHSVLHKSQP-----NVSKG--LVEANLPYTHRIGDSLQSGSPQR 1518
QY 122 -----AHADNVRTKDEINRCK-----QEVKDKNEKYNVS 149
DB 1519 IHNTOFAAEAPOMASLRRVTRTDQHTNRAKLSQQLIEKKSTAYTPTQFGNHSVPMKENS 1578

QY 150 NV-----AVARSQGRYTTNDGCVENP-----ADILEDTGNVAVIVPHGGHYHVPKSDLS 198
DB 1579 NVPSLFAFPAPPKLGTVPNS--YVSPDLVRRESLITSGSEFLP-----PAIGVP 1628
QY 199 ASELAARAKAHLAGNMQPSQLSYSSTPS--PSLPIINFETS--HEKHEEDGYGFDANRIAE 255
DB 1629 TK-----ANSOGSLMYS--PSVEALPIDVVPVPOVHETGYNDFGNKHSQKSMPE 1675
QY 256 DES--GFVMSHGSHNHFFKKDLTEQIKAAOKHLEVKTSNGLDLSLSHEDQYPSNAK 313
DB 1676 DESHTSHDNNAOONTLKSDADVTDEMTDIEGPGFNDVK---NLLPMPNPHQSTVSNPI 1732
QY 314 EMKDLCKIEEKIAGIMQGVKRESIVVKNKNAIYPHGDHHDADPIDEHFKPVGIGHS 373
DB 1733 Q-----TISDDTQPI-LQINVEVRGTDASKMENS-L-PSIENERS---SEEQENISKS 1780
QY 374 HSNVELEKPEGVAKREGNVYVYGEELTNVUNLLKNSFTNNQNFNTLANGOKRVSFSEPP 433
DB 1781 ASS--AYLPSTGGISLE-NRPLTQDE--NSISSETVSTV-----LPAG----- 1818
QY 434 LEKLGINMLVLIITPDGKVLKVGKVGEGVGNIA-----NFELDDQPYLPG 481
DB 1819 -----SISMEAKPLSQVDVPRVNNKASKLVQGHAPPKPKSTDATKNY---SPYVP- 1869
QY 482 QTFKYTIASKDYPEVSDYDGTFTVPTSLAYKMSQ-----TIYPF--HAGDTYLRVNPQ 533
DB 1870 ---QSTAAADGDE-----STILKTSPIAYIARTHQAHASNPQVFFLVNQANET-----AS 1917
QY 534 FAVPKGTDALVRVDFEFGHAYLNNYKVGGEIKLPIPKLN-----QGTTRTAGNK--- 583
DB 1918 FESESTSA-----QSGNNGVASENRF-----PIKAEVVEKDTFOPTIRKASINQYR 1966
QY 584 --IPVTMANAYL-----DNQSTIYIVEFILEKENQTDKPSILPOFKRKAQENSK 632
DB 1967 AFKPLESDADKNDVIEDSDDDNNMSTDEAKNRKEKKVNMKKETKP-----SNKD 2018
QY 633 LDEK-----VEEPKTSKVEKEKLETSNGTSNSTLEEVPTVDVQKVAFAESYGM 685
DB 2019 IDDSNGWFWLKKDQTKGKKVKKHGNLWYDEKLRKWNKDKATEEKKQKLISSAP 2078
QY 686 KLENVLFNMD-----GTIELYLP-----SCEVIKNMADFTGE-----APQG-NG 724
DB 2079 PPPIVKRGKGGPKTKPRSGPINNSLPVHATSVIPNN--PITGEPLPIKTSPTSPTGPNP 2136
QY 725 ENKSENGKVS--TGTVENQFTENKPADSIPEAPNEKP--VKPENSTDNGLNPEGNV 778
DB 2137 NNSPSPSPISRIISGVNLTSSKKNGLDLDLISLAGGPKPASTRRKKTKARGYVNMVMDNI 2194
RESULT 15
MSPL_PIAFM STANDARD: PRT: 1701 AA.
AC P08569;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=70153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.C.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite
RT Plasmodium falciparum";
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP REVISIONS TO 1403; 1569 AND 1629.
RA Tanabe K.;

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:07 ; Search time 37.3021 Seconds

(without alignments)
4639.948 Million cell updates/sec

Title: US-09-471-255-16

Perfect score: 4396

Sequence: 1 CAYALNQHRSQENKONNRVS.....GTIELRLPSGEVKKNLIS 840

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPPREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4274.5	97.2	1039	16 Q9ANY1	Q9any1 streptococc
2	1002.5	22.8	844	2 Q9AG74	Q9ag74 streptococc
3	933	22.6	819	2 Q9ANY3	Q9any3 streptococc
4	989	22.5	819	16 Q97QM9	Q97qm9 streptococc
5	987	22.5	839	16 Q9ANK2	Q9any2 streptococc
6	975.5	22.2	802	16 Q97QM8	Q97qm8 streptococc
7	975.5	22.2	816	2 Q9AHT9	Q9aht9 streptococc
8	867	19.7	825	2 Q93GT5	Q93gt5 streptococc
9	864	19.7	825	16 Q99XV4	Q99xv4 streptococc
10	855.5	19.5	822	2 Q9ZHG7	Q9zHg7 streptococc
11	689	15.7	289	2 Q9AE21	Q9ae21 streptococc
12	224.5	5.1	792	16 Q99Z76	Q99z76 streptococc
13	212.5	4.8	961	16 Q99RD3	Q99rd3 streptococc
14	210.5	4.8	2004	16 Q97QP7	Q97qp7 streptococc
15	208	4.7	1078	5 Q963T1	Q963tl plasmodium
16	196.5	4.5	940	2 Q53682	Q53682 staphylococc

17	194.5	4.4	1964	2 Q59947	Q59947 streptococc
18	194.5	4.4	2485	5 Q96134	Q96134 plasmodium
19	194	4.4	1139	5 Q97237	Q97237 plasmodium
20	191.5	4.4	1939	5 Q25662	Q25662 plasmodium
21	189	4.3	2269	5 Q26223	Q26223 plasmodium
22	189	4.3	2276	2 Q93TV6	Q93ty6 staphylococ
23	188.5	4.3	5458	5 Q9U459	Q9u459 plasmodium
24	187	4.3	1873	2 Q9ZAN7	Q9zan7 enterococcu
25	186	4.2	1524	10 Q8RYN2	Q8ryn2 oryza sativ
26	183	4.2	1236	5 Q9GTX2	Q9gtx2 plasmodium
27	182	4.1	1475	5 Q25842	Q25842 plasmodium
28	180	4.1	1231	2 P72362	P72362 streptococc
29	179	4.1	1650	5 Q77328	Q77328 plasmodium
30	177.5	4.0	1271	5 Q25860	Q25860 plasmodium
31	177	4.0	2747	5 Q9BJX9	Q9bjx9 plasmodium
32	177	4.0	2910	10 Q9FND5	Q9fnd5 arabidopsis
33	176	4.0	2481	16 Q99RG6	Q99rg6 staphylococ
34	175.5	4.0	940	16 Q928N7	Q928n7 listeria in
35	175.5	4.0	1134	2 Q99051	Q99051 streptococc
36	174.5	4.0	1435	5 Q9NG63	Q9ng63 plasmodium
37	173.5	3.9	786	5 Q77357	Q77357 plasmodium
38	173.5	3.9	1141	6 Q46489	Q46489 galago cras
39	173.5	3.9	1576	13 Q9I908	Q9i908 xenopus lae
40	173.5	3.9	1927	2 Q54875	Q54875 streptococc
41	173	3.9	2478	2 Q9RL69	Q9rl69 staphylococ
42	172.5	3.9	3484	5 P91257	P91257 caenorhabdi
43	172	3.9	710	10 Q39060	Q39060 arabidopsis
44	172	3.9	2478	2 Q9LCH2	Q9lch2 staphylococ
45	172	3.9	4688	16 Q9PQ08	Q9pq08 ureaplasma

ALIGNMENTS

RESULT 1

Q9ANY1 PRELIMINARY; PRT; 1039 AA.

AC Q9ANY1;
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
 DE PneuMococcal histidine triad protein E precursor (Hypothetical protein SPI004).
 GN PHE OR SPI004.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21101045; PubMed=11159990;
 RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
 Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
 Langermann S., Koenig S., Johnson S.,
 "Identification and characterization of a novel family of pneumococcal proteins (the Pht family) that are protective against sepsis.",
 Infect. Immun. 69:949-958(2001).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=TIGR4;
 RC MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,
 "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.",
 Science 293:498-506(2001).
 RL EMBL; AF318956; AAK06761.1;

Db 441 KOESLHKLGAARCTDLPSSDREFY-----NKAYDLARIHQDLDLNDKGRQVDEEA- 490
 QY 564 EIKLPIPKLNOGTTRAGNKIPYTWANAYLDNQSYIVEVILEKENOTDKPSILPOFK 623
 Db 491 -----LDNLERLKVSSDKVDLDIAFL-----APIRPE-RLGKPNQIYIT 535
 QY 624 RNKAQNSKL-----DEKVEEPK-TSEK-----VEKEKISSETGNSTSN 660
 Db 536 DDEIQV-AKLAGKYTTEDGYIFDPDITSDEGDYVTPHMTSHHWIKKDSLSEARAAQ 594
 QY 661 STLEEYPTVPQVE-----KYAKFAESY-----GNKLENVLFNMDGIIELYPGE 706
 Db 595 AVAKERGLPTPTDQDSNGTEAKGAEAIYRNVKAAKVPLDRMPINLQYITVE--VKNGS 652
 QY 707 VIKKNMADFTG-----EAPQG-----NGENKPS- 730
 Db 653 LIIPHYDHYHNKEWFDBGLYEAAPKGYLSLELLATVXYVVEHPNPHSDNGFNGASDH 712
 QY 731 -----NGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTNDGMLNPEGVNGSDPMLD 785
 Db 713 VORNNQOQADTQTE-KPNEEKPOTEKPEETPREKQSE-----KPSPKPTPEEP 765
 QY 786 SALEEAP-----AVDPVQEKLEK 803
 Db 766 ESPSPSESEBPQVETEKVKELRE 791

RESULT 3

Q9ANY3 ID Q9ANY3 PRELIMINARY; PRT; 819 AA.
 AC Q9ANY3;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Pneumococcal histidine triad protein B precursor (Fragment).
 GN PHT.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21101045; PubMed=11159990;
 RA Adamcu J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
 RA Dormitzer M., Dagan Y.A., Barren P., Lathigra R.,
 RA Langermann S., Koenig S., Johnson S.;
 RT "Identification and characterization of a novel family of pneumococcal
 RL proteins (the Pht family) that are protective against sepsis."
 RL Infect. Immun. 69:949-958(2001).
 KW EMBL; AF318954; AAK06759.1;
 KW Signal.
 FT SIGNAL
 FT NON_TER
 SQ SEQUENCE 819 AA; 92108 MW; E602FC16CC28A5F CRC64;

Query Match 22.6%; Score 993; DB 2; Length 819;
 Best Local Similarity 31.8%; Pred. No. 9.4e-40;
 Matches 297; Conservative 120; Mismatches 263; Indels 254; Gaps 36;

QY 1 CAYALNQHRS-QENKDNRSYVDGSSQSENITPPQVSKQKGIQAEQIVIKITDQGY 59
 Db 20 CSYELGRYQAGDKKESNRVAYIDGDQAGKAENLTPDEVSKREGINAQIVIKITDQGY 79
 QY 60 VTSBGDHYHYNGKVPYDALFSEELMLKDPNVQLKADIVNEKGGYIKVDGKYVYVLK 119
 Db 80 VTSBGDHYHYNGKVPYDAITISELLMKDPNVQLKDSIVNEIKGGYIVKVGKYVYVLK 139
 QY 120 DAAHADNVTKDEINRQKOEHVKD-NEKVNNSVAVARSQGRYTTNDGVFNPAIDETG 178
 Db 140 DAAHADNIRTKKEIKRQKSHNSHNSRADNAVAAARAGQRYTTDDGVFNASDIIDTG 199
 QY 179 NAYIVPHGGHYHYIPKSDLSASELAAKAHLAKG--NNQPSQLSYSTTSP-----SL 229

Db 200 DAYIVPHGDHYHYIPKNSLASLAAAEAYWNGKQGRPSSSSSYNANPAQPRLESENHL 259
 QY 230 PINGTSH-----EKH-EEDGYGFDANRIIAEDESFGVMHGDNH 268
 Db 260 TVTP-TYHQNOGENISSILLRELYAKPLSERHVESDGLIFDPAQITSRARGVAVPHGNHY 318
 QY 269 HYFFKKDLTEQIKAAOKHLEEV-----KTSHNGLDL-----SSHQDYPSN 311
 Db 319 HF-----IPYQMSLEKRIARIIPLYRSNHWVPDRPEPSPOPTPEPSPPQAPS 373
 QY 312 ---AKEMKDDKKIEEKIAGIMKQYGVKRESIVVNKEKNATIIYPGDDHHDADPIDEKPV 368
 Db 374 PIDGKLKVEARVKVDG--GYVFEENGVS-----YIPAKDLSAETAA 413
 QY 369 GIGHSHSNYELFKPEEGVAKK-----GNKY-----TGEELTN 402
 Db 414 GIDSKLAKQESLSHLGKTKTDLPSDRFYNKAYDALLARIHQDLDLNDKGRQVDFEALDN 473
 QY 403 VVNLKNTFNQNFILANGQKRVSFPPPELEKKLGNMLVKLITPDGKVLKRVSGKF 462
 Db 474 LLERLKOVSSDKVLV---EDILAFAPIRHPERLQKPNNAQIYTTDDEIQVAKLAGKYT 529
 QY 463 GE-----GVGNIANFELDQVLPQGTFKYITASKDYDEVSYDGTFTVPTSLAYKMASOTI 517
 Db 530 AEDGYIFDPDITSDGDAYVTPHMTSHHWIKKDSLSE-----AERAAQA- 575
 QY 518 FYPTHAGDTYLRVNPQFAVPKGTDALVRVDFPHGNAYLENNYKVGKILPIPKLNGOTT 577
 Db 576 YAXEKLIT-----PPSTDHQDSNGTEAKGAEIYRNVKAAK-KVPL----- 615
 QY 578 RTAGNKIPVTMANAYLDNQSYIVEVILEKENQTDKPSILPOFKRNKAQENSKLDEK 637
 Db 616 ---DRMPYNEQ-----YTVEV-----KNGSLIIPHYDHYHNKPEWFEGL 653
 QY 638 -EEPKTEKVEKELSETGNSTNSTLEEVPVDPVQEKVAKFAESYGMKLENVLFNMDG 696
 Db 654 YEAK-----GYTLEDLL-----A 667
 QY 697 TIELYL-----PSGEVIRKKNMADFTGEAPQG-----NGENKPSNGKVGSTGTVENQPTEN 746
 Db 668 TVKYVVEHPNERPHSDNGFGNADHVQRNKGQADTQTEKPSSEKQTEKPEETPRE 727
 QY 747 KPADSLPEAPNEKPV-KPENSTDNGLNPNCGVSDPMLDSALEEAPVDPVQEKLEKFT 805
 Db 728 KPQSEKESP--KPTPEEPSEPE-----SEEPQVET-----EKVEEKLEAE 768
 QY 806 ASYGLGLDSVI-FNMDGTIELRLPSGEVIRKKNL 838
 Db 769 DLLGKIQDPIIKSNAKETL-----TG-LKNLL 795

RESULT 4

Q97QM9 ID Q97QM9 PRELIMINARY; PRT; 819 AA.
 AC Q97QM9;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Conserved domain protein.
 GN SPL1174.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eilen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.F., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Ginn M., Kolonay J.R., Nelson W.C., Peterson J.D.,
 RA Omayun L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,

RA McDonald L.A., Feldblyum T.V., Angiulli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001).
DR EMBL: AE007418; AAK75283.1; -
DR TIGR: SP1174; -
KW Complete proteome.
SQ SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;

Query Match 22.5%; Score 989; DB 16; Length 819;
Best Local Similarity 31.9%; Pred. No. 1.5e-39;
Matches 298; Conservative 118; Mismatches 265; Indels 252; Gaps 36;

QY 1 CAYALNQHRS-QENKDNRRVSYVDGSGSQSKSENITPDQVSKQEGIOAEQIVIKITDQGY 59
DB 20 CSYELGRYQAGQDKKESNRVAYIDGQAGKAENLTDPDEVSKREGINAEQIVIKITDQGY 79
QY 60 VTSHGDHYHYNGKVPYDALFSEELLMKDPNYQLKDADIVNEVKGYIIVKDGKYYVYVK 119
DB 80 VTSHGDHYHYNGKVPYDAIIEELLMKDPNYQLKDSIVNEIKGYIVKNGKYYVYVK 139
QY 120 DAAHADNVRTKDEINRQKOEHVKD-NEKXNSVAVARSQRTTNDGYVFNPAIDIEDTG 178
DB 140 DAAHADNVRTKDEINRQKOEHVKD-NEKXNSVAVARSQRTTNDGYVFNPAIDIEDTG 199
QY 179 NAYIVPHGGHYHYIPKSDLSASELAHAALAGK--NMOPSLSYSTSPSP-----SL 229
DB 200 DAVIVPHGDHYHYIPKNSLSASELAHAAYWNGKGRSPSSSSSYNANPAQPLSENH 259
QY 230 PINPGTSH-----EKH-EDGYGFDANRIITAEDSGFVMSHGHN 268
DB 260 TVTTP-TYHQNQGENISLLRELYAKPLSERHVESDGLIFDPAQITSRARGVAVPHGHY 318
QY 269 HYFFKKDLTEQIAKQAKLEEV-----KTSHNGLDL-----SSHEQDYPSN 311
DB 319 HF-----IPEQNSELEKRIALIPLYRSNHWVDSRPEEPSPOPTPEPSQPAPSN 373
QY 312 AKEMKDLDRKIEKIAAG--IMKQYGVKRESIVYVNEKKNATIIYPHGDHHAADIDEKHPVG 369
DB 374 PIDEK-LVKEAVRKVGDGVFEEGVYSR-----YIPAKDLSAETAAG 414
QY 370 IGHSHSNYELFKPEEGVAKKE-----GNKY-----TGBELTNV 403
DB 415 IDSKLAKQESLSKLGKTKKTDLPSSDRBYNKAYDILLARIHQDILLDNKGRQVDFEALNL 474
QY 404 VNLKRNSTFNQNTFLANGQKRVFSFPPELEKKLGINMLVLTDPDGKVLKVGSKVFG 463
DB 475 LERLKDVSDDKVLV---EDILAFAPIRHPERLKGPNQAITYTDEIQVAKLAGKYTT 530
QY 464 E-----GVGNIAINFELDQPLPQOTKYTKYASKDYPEVSDGTFTVPTSLAYKMASOTIF 518
DB 531 EDGYIFEDPRDITSDGDAYVTPHMTSHWIKKDSLSE-----AERAAQA--- 575
QY 519 YPFHAGTYLVRNPOFAVPKGTDALVRVDFEFGNAYLENKYKVEIKLPIPKLNOGTT 578
DB 576 YAKEKGLT-----PSTDHQDSGNTAKGAEAIRNKAAR-KVEL----- 615
QY 579 TAGNKPVTMANAYLDNQSTYIIVEPILKENQTDKPSILPQFKRNKAQENSKLDEKV- 637
DB 616 ---DRMPYNLQ-----YITVEV-----KNGSLIIPHIDYHNIKFEWFDEGLY 654
QY 638 EEPKTSKVEKEKLSERGNSTNSLFEVPTVDPQEKVAKFAESYCMKLENVLFNMDGT 697
DB 655 EAPK-----GYTLEDLL-----AT 668
QY 698 IEYL-----PSGEVTKKNWADFTGEAPG-----NENKPSNGKVSIGTVENQPTENK 747
DB 669 VKYVVEHPNERPHSDNGFNASDHVRQNKNGQADTNTQTEKPSSEKQPTKEPEETPREK 728
QY 748 PADSLPAPNEKPV-KPENSTDNGLNPNPEGVSDGMDLSALEAPAVDPQEKLEKFTA 806

DB 729 PQSEKPESP-KPTEPEESPESPEE-----SEEPQVET-----EKVEKLREARD 769
QY 807 SYGLGLDSVI-FNMDGTIELRLPSGEVIRKNNLL 838
DB 770 LLGKIQDPIKSNKAKETL-----TG-LKNNLL 795

RESULT 5
Q9ANY2
ID Q9ANY2 PRELIMINARY: PRT: 839 AA.
AC Q9ANY2;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE Pneumococcal histidine triad protein B precursor (Hypothetical protein
DE SP1003) (Fragment).
GN PHTD OR SP1003.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBL_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.,
RA "Identification and characterization of a novel family of pneumococcal
RT proteins (the pht family) that are protective against sepsis";
RL Infect. Immun. 69:949-958(2001).
RN [2]
RP SEQUENCE FROM N.A.
PC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn O., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiulli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506(2001).
DR EMBL: AF318955; AAK06760.1; -
DR EMBL: AE007403; AAK75120.1; -
DR TIGR: SP1003; -
KW Signal; Hypothetical protein; Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 839 839
SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;

Query Match 22.5%; Score 987; DB 16; Length 839;
Best Local Similarity 32.08; Pred. No. 1.9e-39;
Matches 299; Conservative 106; Mismatches 278; Indels 252; Gaps 33;

QY 1 CAYALNQHRS-QENKDNRRVSYVDGSGSQSKSENITPDQVSKQEGIOAEQIVIKITDQGY 59
DB 20 CSYELGRYQAGQDKKESNRVAYIDGQAGKAENLTDPDEVSKREGINAEQIVIKITDQGY 79
QY 60 VTSHGDHYHYNGKVPYDALFSEELLMKDPNYQLKDADIVNEVKGYIIVKDGKYYVYVK 119
DB 80 VTSHGDHYHYNGKVPYDAIIEELLMKDPNYQLKDSIVNEIKGYIVKNGKYYVYVK 139
QY 120 DAAHADNVRTKDEINRQKOEHVKDNEKXNSN---VAVARSQRTTNDGYVFNPAIDIED 176
DB 140 DAAHADNVRTKDEINRQKOEHVKDNEKXNSN---VAVARSQRTTNDGYVFNPAIDIED 198
QY 177 TGNAYIVPHGGHYHYIPKSDLSASELAHAALAGK--NMOPSLSYSTSPSP----- 227
DB 199 TGDAYIVPHGDHYHYIPKNSLSASELAHAAYWNGKGRSPSSSSSYNANPAQPLSENH 258

AC Q9AHT9: 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Pneumococcal histidine triad A protein.
GN PHTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RX MEDLINE=21116976; PubMed=11179332;
RA Wisemann T.M., Heinrichs J.H., Adomou J.E., Erwin A.L., Kunsch C.,
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
RA Langermann S., Johnson S., Koenig S.,
RT "Use of a Whole Genome Approach To Identify Vaccine Molecules
Affording Protection against Streptococcus pneumoniae Infection.",
RL Infect. Immun. 69:1593-1598(2001).
DR EMBL: AF291695; AAK39155.1;
SQ SEQUENCE 816 AA; 91319 MW; 5359126A611D27ED CRC64;

Query Match 22.2%; Score 975.5; DB 2; Length 816;
Best Local Similarity 33.9%; Pred. No. 6.4e-39;
Matches 271; Conservative 105; Mismatches 203; Indels 221; Gaps 29;

QY 1 CAYALNHRSQENKNNRVSYVDGSSQSKSENLTDPDOVSQKEGIAEQIVIKITDQGYV 60
DB 20 CSYELGLQARTVKENRVSYIDGQATQKTENLTDPDEVSREGNAEQIVIKITDQGYV 79
QY 61 TSGHDHYHYNGKVPYDALFSEELLMKDPNQLKADAVNEVKGYYIIKVDGKYYVYLK 120
DB 80 TSGHDHYHYNGKVPYDALFSEELLMKDPNQLKADAVNEVKGYYIIKVDGKYYVYLK 139
QY 121 AAHADNVRTKDEINRQKQHVKDNE----KYNVNVAARSQGRITNDGYVFNPAIDIED 176
DB 140 AAHADNVRTKEEINRQKQHVKGREGGTPRNDGAVALARSGRYTTDDGYIFNASDIED 199
QY 177 TGNAYIVPHGGHYHYTPKSDLSASLAALAAKHLAGKNNQPSOLSYSSTPS-----P 227
DB 200 TGDYIVPHGDHYHYTPKNELSASELAALAAFLSGRGLNSRTPYRQNSDNTSRTNVP 259
QY 228 SLPIPGTSE-----EKH-EDGVGFANDRIIA 254
DB 260 SVS-NGPTTNTNTSNTNSQASQSDNIDSLKLYKLPLSRHVESDGLVFDPAQITS 318
QY 255 EDSGFMVSHGDHNYFFKDLTEQIKAAQKHLEBVKTSNGLDLSLSHEDYD----- 309
DB 319 RTARGVAVPHGDHYHPIPYQSQSELEERTARIIPLYRSNHWVPSRPEQSPPTPEPS 378
QY 310 -----SNAKEMKDLKKIEEKIAGIMKQYGVK-----ESIVNKEKNAI 349
DB 379 PGQPAPNLKIDNSLSVLQVRKVE--GYVEEKGISRYVFAKDLPSRTVKNLESKL- 435
QY 350 IYPHGDHHDADPIDENKPVGIGHSHSNYELFPEEGVAKKEG--NKVY---TGEELNV 403
DB 436 -----SKQSVSH-----LTAKKNVAPROEFTDKAYNLLTEAHKALF 475
QY 404 VNLLKNSTF-----NNQFTLANGKRV-----SFSPPELEKLGINMLVLTITPGK 452
DB 476 XNKGNSDFQALDKLLERLNDSTNKEKLVDDLLAFLAPITHTPERLG-----KPSNQ 527
QY 453 VLEKVSQGVFEGGVNIAFELDQPLPQCTFKYTIAS-----KDYPEVSDGTFVPSLSL 508
DB 528 I-----EYTEDVIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHM 570
QY 509 AY-----KMASQTFYPPHAGDTYLRVNPQFAVPKGTDALVRVDFEFHGNAY-555
DB 571 GSHHWTKGKSLDSKEKVAQA--YTKEG-----ILPSPDADVKANPTGDSRAA 618
QY 556 LENNYKVGKELPKLPKNOGTTTAA-----GNKI-----PVTMANAYLDNQS-----598

DB 619 IYNRVK-GEKRIPLVRLPYMVEHTVEVKNGLIIPHKDHYHNKIF---AWEDDHTYKAPN 674
QY 599 -----TYIVEVP-----ILEKENOTDKPSILPQFKRNKAQNS 631
DB 675 GYTLEDLFATIKYIVVEHPDERPHSDNGWGNASEHVLGKKHSDSDPN-----KNFADE-E 728
QY 632 KLDEKVEEPTSEKVEKEL 651
DB 729 PVETAEPEVPQ-VETEKV 747

RESULT 8
Q93GT5 PRELIMINARY; PRT; 825 AA.
ID Q93GT5;
AC Q93GT5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Histidine triad protein of group A streptococci.
GN HTPA.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-9;
RA Terao Y., Kawabata S., Hamada S.,
RT "Characterization of a novel histidine triad protein of group A
streptococci.",
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB073859; BAB7774.1;
SQ SEQUENCE 825 AA; 92623 MW; DE4ECC199181DFFB CRC64;

Query Match 19.7%; Score 867; DB 2; Length 825;
Best Local Similarity 28.4%; Pred. No. 1e-33;
Matches 267; Conservative 145; Mismatches 272; Indels 256; Gaps 43;

QY 2 AYALNQRSQENKNNRVSYVDGSSQSKS--ENLTPDOVSQKEGIAEQIVIKITDQGY 59
DB 22 SYQLGKHMGMSATKDNQIAYIDDSKGRKAPKTNKMDQISAEEGISAEOIVVKTIDQGY 81
QY 60 VTSGHDHYHYNGKVPYDALFSEELLMKDPNQLKADAVNEVKGYYIIKVDGKYYVYLK 119
DB 82 VTSGHDHYHYNGKVPYDALFSEELLMTDNYREKQSDVINEILDGVIVKNGNYVYLK 141
QY 120 DAHADNVRTKDEINRQKQHVKD-NEKVNINVA-----VARSGRITNDGYV 167
DB 142 PGSKRNIRTKQQAQVAKGTKEAKEGLAQVAHLSKEEVAAVNEAKRGRTYTDGDI 201
QY 168 FNPADITEDTGNAYIVPHGGHYHYTPKSDLSASLAALAAKHLA---CKNMQPSQLSYSST 224
DB 202 FSPIDIIDLDGDAYLVPHGNYHYIIPKDLSPSELAAQAYWSQKQGRGAPS--DYRPT 259
QY 225 PPS-----LPI-----NPGTSH-----KH----- 240
DB 260 PAPAGRRKAPIDVTPNPGQHQPDNGVYHPPRPNDASQNKHQDEKGTFKELLD 319
QY 241 -----EDGVGFANDRIIAEDESQFMVSHGDHNYFFKDLTEQIKAAQKH 288
DB 320 QLRDLKRYRHVEDGLIFEPTQIKSNAGVYVPHGDHIIIPRSQSLPELELADRY- 378
QY 289 EEVKTSHNGLDLSLSHEDQYPSNAKEMKDLKKIEEKIAG-IMKQYGV-----KRES 339
DB 379 -----LAGQEDDDSGDSKSPSKVEVTHFLGHRIKAYKGLDGKPYDTSDA 426
QY 340 IVYNKE-----KNAIYPHGDHHDADPIDENKPVGIGHSHSNYELFPEEGVAKKEG 393
DB 427 YVFSKESTHSYDKSGVTAKGDHFIY-----IGFGELEQELDEVANWV-KAQG- 475
QY 394 VYTGELTNVYLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINMLVLTITPGKY 453

```
Db 476 ---ADELAALD-----QEQEKPLFD-----TKKVS-----RKVTGDKGV 509
QY 454 --LEKVSQVFGVGVGNIANFELDQYLPQGTF-KYTIASKDYPEVSYDGTFTVPTSLAY 510
Db 510 GYMMPKDGKY-----FYARDQLDLTQIAFAOELMLKDKKHRYD---IVDTGTGP 558
QY 511 KMASQTIFFPFIAGDTYLRVNPQFAPVPKGTDALRVDFDEHGNAYLNNYKVEIKLPTIP 570
Db 559 RLAVDVSSLPFHAGNATYDGTGSSFVPH-----IDHIH----- 591
QY 571 KLNQGTTRTAGNKIPVTFMANAYLDNQST--YIVEVP-----ILEKNGTDKPSILPQFK 623
Db 592 -----VVYPSWLTR---DOIATIKYVMQHPVPRDINSKPGHEESGVIPNV- 635
QY 624 RNKAQNSKLDKEVEEPK-----TSEKVEKEKLSGTNSTSLTEEEVTPVDPQEKVAK 678
Db 636 -----TPLDKRAGMPNQIHSAEVQK-ALAEGRFATPDGYI-----FDP-RDVLAK 681
QY 679 FAESYGMKLENVLFNMDGTIELYLPSEGVIKK-NNAFTGEAPQNGENKPSGVSTG 737
Db 682 ---ETFWVK-----DGSFSPRADGSSLRITINKSDLSQAEWQOQALLAKKNAGDATD 731
QY 738 TVENQPTENKPADSLPE--APNEKVPKPNSTDNMGNLPEGNVGDPMDSALAEAPAVD 795
Db 732 T--DKPKQKQADKSNENQPSAEKKEESDDFI-----DSLDPYGLDRATLED 780
QY 796 PVQEKLEKTASYGLGLDS--VIFNMDGTIELRLPSGEVI 833
Db 781 HINLAQK-----ANIDPKYLIFQPEG-VQFYNNKNGELV 813

RESULT 9
Q99XV4
ID Q99XV4 PRELIMINARY; PRT: 825 AA.
AC Q99XV4:
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical protein SPY2006.
GN SPY2006
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11298296;
RA Ferrer J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: AE006623; AAK34688.1; .
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 825 AA; 92649 MW; 57DF50969E6D50F4 CRC64;

Query Match 19.7%; Score 864; DB 16; Length 825;
Best Local Similarity 28.4%; Pred. No. 1, 4e-33;
Matches 267; Conservative 144; Mismatches 273; Indels 256; Gaps 43;

QY 2 AYALNQHRSQENKNNRVSVDGSSQSKS--ENLTPQVQKEGIAEQIVIKITDQGY 59
Db 22 SYQLGKHGMSATKDNQAIYIDDSKGAKAPKTKMTDQISAEGISAEQIVWKITDQGY 81
QY 60 VTSHGDDHYHYNGKVPYDALFSEILLMDPNYQLKADIVNEKGYLIKVDGKYYVYLK 119
Db 82 VTSHGDDHYHYNGKVPYDAIISEELLMTDPNRFKOSDVINEILDGYIKYNGNYYVYLK 141
QY 120 DAAHADNVRITKDEINROKQEHVKD-NEKVNNSVA-----VARSQGYTTNDGYV 167
Db 142 PGSKRKNIRTKQIAEQVAKGFKKEAKERGLAQVLAKEEVAANVEAKRQGYTTDGYI 201
```

```
QY 168 FNPADIIEDTGNAYTPHGHHYHYPKSDLSASELAACAHLA---GKNMOPSQLSYSS 224
Db 202 FSPDTIIDLDGAILVPHGNHYHHPKDKLSPSELAACAQAIYKQSGRGARPS---DIRPT 259
QY 225 PSPS-----LPT-----NPGTSHE-----KH----- 240
Db 260 PAPAPGRRKAPIDPVTNPNGOGHGDNGGYHPAPRPNDASONKHORDEFKGTKEKLLD 319
QY 241 -----EEDGYCFDANRIIAEDESGFVMSHGDHNNHYFFKDLTEQIYKAAQKHL 288
Db 320 QLHRLDLKYRVEDGLIFEPTQVIKSNAGYVYPHGDDHYHIIIPRSQLSPLEMLADRY- 378
QY 289 BEVKTSHNGLSLSSHEODYPSNAKEMDKDKTEEKIAG-IMQYGV-----KRES 339
Db 379 -----LAGOTEDDSDGSDHSPKSPSKVETHIFLGHRIKAYKGLDKGKYDTSDA 426
QY 340 IIVNKE-----KNAIYPHGDHHDADIDHKPVGCHSHSNYELFKPPEGVAKKGNK 393
Db 427 YVFSKESITHSVKSGVTAHGDHFHY-----IGFEL-EQYELDEVANWV-KARQ- 475
QY 394 VYTGEELTNVNLKNSFTNQNTLANGOKRVSFSPPELEKKLGINMLVKLITPDGKV 453
Db 476 ---ADELAALD-----QEQEKPLFD-----TKKVS-----RKVTGDKGV 509
QY 454 --LEKVSQVFGVGVGNIANFELDQYLPQGTF-KYTIASKDYPEVSYDGTFTVPTSLAY 510
Db 510 GYMMPKDGKY-----FYARDQLDLTQIAFAOELMLKDKKHRYD---IVDTGTGP 558
QY 511 KMASQTIFFPFIAGDTYLRVNPQFAPVPKGTDALRVDFDEHGNAYLNNYKVEIKLPI 570
Db 559 RLAVDVSSLPFHAGNATYDGTGSSFVPH-----IDHIH----- 591
QY 571 KLNQGTTRTAGNKIPVTFMANAYLDNQST--YIVEVP-----ILEKNGTDKPSILPQFK 623
Db 592 -----VVYPSWLTR---DOIATIKYVMQHPVPRDINSKPGHEESGVIPNV- 635
QY 624 RNKAQNSKLDKEVEEPK-----TSEKVEKEKLSGTNSTSLTEEEVTPVDPQEKVAK 678
Db 636 -----TPLDKRAGMPNQIHSAEVQK-ALAEGRFATPDGYI-----FDP-RDVLAK 681
QY 679 FAESYGMKLENVLFNMDGTIELYLPSEGVIKK-NNAFTGEAPQNGENKPSGVSTG 737
Db 682 ---ETFWVK-----DGSFSPRADGSSLRITINKSDLSQAEWQOQALLAKKNAGDATD 731
QY 738 TVENQPTENKPADSLPE--APNEKVPKPNSTDNMGNLPEGNVGDPMDSALAEAPAVD 795
Db 732 T--DKPKQKQADKSNENQPSAEKKEESDDFI-----DSLDPYGLDRATLED 780
QY 796 PVQEKLEKTASYGLGLDS--VIFNMDGTIELRLPSGEVI 833
Db 781 HINLAQK-----ANIDPKYLIFQPEG-VQFYNNKNGELV 813

RESULT 10
Q92HG7
ID Q92HG7 PRELIMINARY; PRT: 822 AA.
AC Q92HG7:
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical 92.4 kDa protein.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R268;
RX MEDLINE=99115568; PubMed=9916102;
RA Spellerberg B., Rozdzinski E., Martin S., Weber-Heymann J.,
RA Schnitzler N., Luetticken R., Podbielski A.;
RT "Lmb, a protein with similarities to the Lrai adhesin family, mediates
```

RT attachment of Streptococcus agalactiae to human laminin.";

RL Infect. Immun. 67:871-878(1999).

DR EMBL; AF062533; AAD13797.1; .

KW Hypothetical protein.

SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;

```

Query Match          19.5%; Score 855.5; DB 2; Length 822;
Best Local Similarity 29.0%; Pred. No. 3.6e-33;
Matches 270; Conservative 143; Mismatches 277; Indels 241; Gaps 45;

QY 2 AYALNQRHSQENKDNRRVSVDSQSQSKS--ENLTPDQVSQKEGIAEQIIVKIDTQGY 59
DB 22 SYOLGKHMGKATKDNQIAYIDDSKGVKAPKTKNTMDQISAEEGISAEQIIVKIDTQGY 81
QY 60 VTSHGCHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGYIIVKIDGKYVYLK 119
DB 82 VTSHGCHYHYNGKVPYDAIISELLMTDPNHFQKSDVINEILDGVIIVKNGNYVYLK 141
QY 120 DAHADNVRTKDEINRQKQEHVKD-NEKVNNSVA-----VARSGRYTTNDGYV 167
DB 142 PGSKRNIKTQKQIAEQVAKGTKEAKEGLAQVAHLSKEEVAAVNEAKRGYTTDDGYI 201
QY 168 FNPADIIEDTGNAYIVPHGHHYIIPKSDLSASELAHAKAHLA---GKNMQPSQLSYST 224
DB 202 FSPTDIIDDLDGAYLVPHGNHNYIIPKDLSPSELAQAQYWSQKQGRARPS--DYRPT 259
QY 225 PSP---SLPT-----NPGTSHE-----KH----- 240
DB 260 PAPGRKAPIDPVTNPGQGHQDPNGGYHPAPPNPNDASQNHQRFKGTFFKELLDL 319
QY 241 -----EEDGYFDANRIIAEDSGFVMSHGDNHNYFFKDLTEQIKAAQKHLEE 290
DB 320 HRLDLKYRHEVEDGLIFPTQVTKSNAGFYVPHGHHYIIPRSQLSPLEMLADRYLAG 379
QY 291 VKTSHNGCLDSLSEHQDYPSNAK-----ENKOLDKIEEKIAGIMQYGVKRESIVV 342
DB 380 -QDDN--DSGSDHSK--PSDKVETWTF.LGRIKAYKGGLDGKPYDTSAYVFSKESI-H 433
QY 343 NKEKNALIYPHGHGHHADPIDIEHKPKYIGIHSNYSYLFPEGVAKKEGKNKYVTGELTN 402
DB 434 SVDSKGTAKEGDFHY-----IGFGL-BOYELDEVANWV-KAKGO-----ADELVA 479
QY 403 VVLLKNSTNNQNTFLANGOKRVYSFPPELEKLGIMLVKLITPDGKV--LEKVSCK 460
DB 480 ALD-----QEOGKERPLED-----TKKVS-----RKVTGDKGVYIMPKDCK 516
QY 461 VFGEGVGNIANFELDQPLPFGQFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYP 520
DB 517 DY-----FYARYQLDLTOI--AFAGQELMLKDKKHRYD--IVDTGIEPLRAVDVSSLP 566
QY 521 FHAGDTYLRVNPQFAPVPGKTDALVRVDFEFGNAY--LENNYKVGEIKLPIPKLNOGTR 578
DB 567 MHAGNATPYTGSSSFVIPH-----IDHIVVYPSWLTRN-OIATIK----- 605
QY 579 TAGNKIPVTMANAYLDNQSTYIIVEP-----ILEXENOTDKPSILPQFKRNKAQENSKL 633
DB 606 -----YVQHPVEPVPDWSRPGHESGSVIPNV-----TPL 636
QY 634 DEKVEEPK-----TSEKVEKEKLESETGNSTNSLTLEVPVTPDQVKVAKFAESYGMKLE 688
DB 637 DKRAGMNPWQIISAEVQK-ALAEGRFAAPDGYI-----FDP-RDVLAK--ETFWVK-- 685
QY 689 NVLFNMDGTTLELIPSEVTKK-NMADFTCEAPQNGENKPSNGKVSIGCTVENQPTENK 747
DB 686 -----DGSFSIPRADGSSURTINKSDLSAQEQACQALLAKKNAGDADTD--DKPEEKQ 737
QY 748 PADSLPAPNEKP---VKPNSNDSTNGMLNPEGNVGDPMPLDSALEAPAVDPVQEKLEKF 804
DB 738 QADKSNE--NQQPSSEASKEEKESDDFI-----DSLDPDYGLDRATLEDHINQLAQK- 785
QY 805 TASYGLGLDS--VIFNMDGIIELRLPSGEVI 833
DB 786 -----ANIDPKYLIFQPEG-VQFYNKNGELV 810

```

RESULT 11

ID Q9AE21 PRELIMINARY; PRT; 289 AA.

AC Q9AE21;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Hypothetical 32.0 kDa protein (Fragment).

OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1311;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=5531;

RA MEDLINE=21172873; PubMed=11274116;

RX Granlund M., Michel F., Norgren M.;

RT "Mutually exclusive distribution of IS1548 and GBSII, an active group

II intron identified in human isolates of group b streptococci.";

PL J. Bacteriol. 183:2560-2569(2001).

DR EMBL; AJ290952; CAC35985.1; .

KW Hypothetical protein.

FT NON_TER 289 289

SQ SEQUENCE 289 AA; 32043 MW; A15A8588EA8140E4 CRC64;

Query Match 15.7%; Score 689; DB 2; Length 289;

Best Local Similarity 51.3%; Pred. No. 8.3e-26;

Matches 138; Conservative 42; Mismatches 61; Indels 28; Gaps 8;

QY 2 AYALNQRHSQENKDNRRVSVDSQSQSKS--ENLTPDQVSQKEGIAEQIIVKIDTQGY 59

DB 22 SYOLGKHMGKATKDNQIAYIDDSKGVKAPKTKNTMDQISAEEGISAEQIIVKIDTQGY 81

QY 60 VTSHGCHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGYIIVKIDGKYVYLK 119

DB 82 VTSHGCHYHYNGKVPYDAIISELLMTDPNHFQKSDVINEILDGVIIVKNGNYVYLK 141

QY 120 DAHADNVRTKDEINRQKQEHVKD-NEKVNNSVA-----VARSGRYTTNDGYV 167

DB 142 PGSKRNIKTQKQIAEQVAKGTKEAKEGLAQVAHLSKEEVAAVNEAKRGYTTDDGYI 201

QY 168 FNPADIIEDTGNAYIVPHGHHYIIPKSDLSASELAHAKAHLA---GKNMQPSQLSYST 224

DB 202 FSPTDIIDDLDGAYLVPHGNHNYIIPKDLSPSELAQAQYWSQKQGRARPS--DYRPT 259

QY 225 PSP---SLPT-----NPGTSHEKEEDGY 245

DB 260 PAPGRKAPIDPVTNPGQGHQ-PDNGGY 287

RESULT 12

ID Q99Z76 PRELIMINARY; PRT; 792 AA.

AC Q99Z76;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Putative internalin A precursor.

GN INLA OR SPY1361.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;

OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1314;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;

RA MEDLINE=21192684; PubMed=11296296;

RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,

Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,

Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,

Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;

QY	755 A 755
DB	746 A 746
RESULT 13	
Q99RD3	PRELIMINARY; PRT: 961 AA.
ID	Q99RD3
AC	AQ9RD3;
DT	01-JUN-2001 (TrEMBLrel_17, Created)
DT	01-JUN-2001 (TrEMBLrel_17, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel_21, Last annotation update)
DE	FnbB protein Fibronectin-binding protein homolog).
GN	FNBB OR SAV2502 OR SA2290.
OS	Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC	Staphylococcus aureus (strain N315).
OC	Bacteria; firmicutes; Bacillus/Clostridium group; Bacillales;
OC	Staphylococcus
OX	NCBI_TaxID=158878; 158879;
[1]	
RN	SEQUENCE FROM N.A.
RC	SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RC	MEDLINE=21311952; PubMed=11418146;
RA	Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA	Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA	Kanamori M., Matsumaru R.H., Maruyama A., Murakami H., Hosoyama A.,
RA	Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA	Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA	Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.
RA	Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.
RT	"Whole genome sequencing of methicillin-resistant Staphylococcus
RT	aureus".
RL	Lancet. 357:1225-1240(2001).
DR	ENBL; AF003365; BAB58664.1; -
DR	EMBL; AF003137; BAB43593.1; -
DR	InterPro: IPR004237; Fm_bind.
DR	InterPro: IPR001899; Gram_pos_anchor.
DR	InterPro: IPR002965; P-rich_extensn.
DR	Pfam; PF02986; Fm_bind; 1..anchor;
DR	Pfam; PF00746; Gram_pos_anchor; 1.
DR	PRINTS; PR01217; PRICHEXTENS.
DR	TIGRFAMS; TIGR01167; LPXTG_anchor; 1.
DR	TIGRFAMS; TIGR01168; YSRK_signal; 1.
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW	Complete proteome.
SEQ	SEQUENCE 961 AA; 106010 MW; 364940F884EA4101 CRC64;
Query Match	4.8%; Score 212.5; DB 16; Length 961;
Best Local similarity	21.4%; Pred. No. 0.029;
Matches	17%; Conservative 110; Mismatches 287; Indels 251; Gaps
QY	26 QSSOKSE-NLTTPDVSOKEGIOAEIVIKITDGGVYTSHG---DHVHYNKVPYPDALPSE 82
DB	: : : : : : : : : : : : : :
DB	290 QGSQKEVEVLGQNKKVKSFEDI---KYLDGVKDMRGVTVMNGRIDTLNKEGK-----FSH 340
QY	83 ELAKMDPNYQLKDADIVNEKVGGYIIKVDG---KYVVYLKDAHADNVRTKDEINRKQKF 139
DB	: : : : : : : : : : : : : :
DB	341 FAXYVKNNNSLTSVTGTGQSYGKSANNPTVKYKHIGSDLAESAAYVAKLD-DTSKFE 399
QY	140 HVKDNKKNVSNVAARSQGRYTTNDGYVENPADIEDTCNAYIVPGHGHHYIPKSDLSA 199
DB	: : : : : : : : : : : : : :
DB	400 DV---TEKVNLS-----YTSNGYTLNLGDL---DNSKDVIKYEKG-----DQA 440
QY	200 SELAAKAHLAGLNMGPSQLYSSTSPSLPINPGTSHKHEEDYGGFDA--NRIIAEE 257
DB	: : : : : : : : : : : : : :
DB	441 KDL-NERTHLSGHVHKYPYPPYPPYPPVOLTNNGVAFYSNNAKGDGDKDPNPITEKSE 499
QY	258 SGFMVSGHDHNHFYFKDTEEQIKAQKH-----LEEVKTS-----HNGDSLSSHE 305
DB	: : : : : : : : : : : : : :
DB	500 -----PIDDIKSEPPEVKHELTGITIEESNDSKPIDFYFHTTAVEAGEHA 544
QY	306 ODVPSNAKENKDDUKITEFKIAGIMKOYGVKFSVIVNPKFAIVTPHGDTHHHPIDPH 365

```

Db 545 E-----GIIIE-----TEEDSIHVDFEEST-----HENSXKHADVVEYE 577
QY 366 KPVGIG-----HSHSNYELFKPE-EGV-----AKKCKGNKYVTGEELTNVNL-----406
Db 578 EDTNPGGGQVVTESNLVFEDEESTKGIIVTCAVSDHTIVETDKYITE--SNLIELVDLDP 635
QY 407 -----LNKNFTNNQNT-----LANGQRVSFSFPPPELEKKGILNMLVKLITPDCKVL 454
Db 636 BEHGCAQGPPIEITENNHHISHSLGTENGHNGYGVIDEIENSHRVDIKSEL-----687
QY 455 EKVSQKVPFGGVGNIAFELDQVLPQGTFTYIAKQKXPEVSDGTFTVPTSLAYKMAS 514
Db 688 -----GYEGQNSGN-QSFEED-----TEDRKPYEQGNI-----717
QY 515 QTIFYPFHAGDTYLRVNPQFAVPGKTDALRVDFEFHGNAYLNNYKVGELKIPKLNQ 574
Db 718 -----VDIDE-----DSVPQIHQNGNGQSFEDTEEDK-----PKYEQ 751
QY 575 GTTRTAGNKIPVTMANYLDNSTYIVVPILKEKNOTDKPSILQFKRKAQENSKLD 634
Db 752 G-----GNLIDIDFSVPQIHGNKH-----NEIIEEDTKDKPNY--QF-----GGRNSVD 797
QY 635 EKVEEPTSEKVEREKLETSNGTSTNSLTLEVP-----VDPVQEKVAKFAESYGMKLEN 689
Db 798 EEDTLKVSQNGQEQOITEEDTTPPTPTPEVPSEPTPTPTPEVPSEPTPTPKPE- 856
QY 690 VLFNMDGTIELYLSGSEVIVKKNADFTGEAPQNGE-----NKPSNGKVSSTG 737
Db 857 -----VPSEPTPVPP-----TPVPSGPKVPVPAKEEPKPKPVQGGKVVTP 901
QY 738 TVE-NQ-----PTENKPA--DSLPEAPNEKPKVPKPNSTNDGML 772
Db 902 VIEINERYKAVAPTQKQSKKSELPETGEE-----STNKGML 939

RESULT 14
Q97Q7 ID Q97Q7 PRELIMINARY; PRT: 2004 AA.
AC Q97Q7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Immunoglobulin A1 protease.
CN Sp1154.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TI6R4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tectelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Keldibiyum T.V., Anghuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007416; AAK75263.1;
DR MEROPS; M26.001;
DR TIGR; Sp1154;
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001130; Zn_MTPptdse.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR TIGRFAMs; TIGR01168; YSIRK_signal; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

```

```

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease; Complete proteome.
SQ SEQUENCE 2004 AA; 223908 MW; 556BC6A1028D60AA CRC64;

Query Match 4.8%; Score 210.5; DB 16; Length 2004;
Best Local Similarity 22.1%; Pred. No. 0.098;
Matches 17; Conservative 93; Mismatches 241; Indels 295; Gaps 47;

QY 186 GGHVYIPKSDLSASELAARAKHLACKNMOPSLSY-----SSPT 225
Db 47 GVHYKYVADSLSEE-----KKQLVYDIPTYVENDDETYLVYKLNQ 92
QY 226 SPSLPIPIPGTISHEKHE--EDGYGFDANRIIAEDSGVMSGHGHNHFFKDKTEQIKAA 284
Db 93 LAELP-NTGSKNEHQALVAGASLAAMGILI-----FAVS-----KKVKNKT 134
QY 285 QKHEEYKTSNG-----LDSLSH-----BQYPSNAKEMKDLKRIE-EKTAGINKQY 334
Db 135 -LHLVLVAGTNGVLSVSHALENHLNLYNTDYELTSGEKLPLPKETISGTYVIGYKEGK 193
QY 335 VKRESIVNNEKNAIIVP-----HGDH-----HHADPIDEHPVIG 371
Db 194 TTSSEVSN-QKSSVATPTKQKVYVNTFVDHPSVQAIQEQTPVSTKPTVEQVVE 252
QY 372 -----HSHSNYELFKPEGVAKREGKRYT-----GEELTNV 403
Db 253 KPFSTELINPRKEKQSSDSQEOIAEHKNET-KKEKISPKETGVNTLNQDEVLSQ 311
QY 404 VN-----LLKNSTFN-----NQNTLANGOKRVSFSPPELEKKGILNMLVKLIT 448
Db 312 LNKPELLYREETMETKIDFQETCEPNDLAEGTVR-----KQEGKLGKKEIVRIEVS 365
QY 449 -----PDGKLVKVSCK--VFGE-----GV-----GNIAFELDQVYL 479
Db 366 NKEEVSREIVSTSTTAPSPRIVEKTKTKTQVIKEQPTGVHEKDVQSGAIVPAL-QPEL 424
QY 480 PGQTFKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIYPPHAGDIYLRVNPQ-----FA 535
Db 425 P-----EAVSDKGEPEVQ-----TLPEAV-----TDKGET--EVQPESPDVV 463
QY 536 VPKGTDALRVDFEFHGNAYLNNYKVGELK--LPIPKL-NQOTTRTAGNKIPVTFMANA 592
Db 464 SDKGEPOQVAPLPEYKGN-----IEQVKPTPEKTKEQGPEKT--EEVPV-----507
QY 593 YLDNQSTYIIVVPILKEKNOTDKPSILQPKRN---KAQENSKLDEKVEEPTSEK-----645
Db 508 -----KPTETPVNPNEGTGTSTI--QDAENPVQPAEESTTNSSEK-SPOTSSKNWGE 558
QY 646 -----VEKEKLSSETGNTSTNLTSEVPTVDPVQEKVAKFAESYGMKLENYL 691
Db 559 VSSNPSDSTTSVGSBKNPKHNSKENSEKTVVEVP-VNP-----597
QY 692 FNMGTIELYLPSCGEVILKNWADFTGEAPQNGENKPESENGKV---STGTVENQPTENKP 748
Db 598 --NGSTVE-----GTSNQETKPVQPAETQTSNGKANTANETGEVSNKPSDSKP 644
QY 749 --ADSLPEAPNEKPKVPKENS-----TDNGMLNPE---GNVGSDFMLDSALEEAPVDPVQE 799
Db 645 PVESNQPEKNGATKATKPENSGNTTSENGOTEPEPSNGNSTEDVSTESNTSNGNEELIQ 704
QY 800 KLEKFTASYGLGLDSVIFNMDGIELR 826
Db 705 ENE-----LDPPDKVVEEPEKTELEUR 724

RESULT 15
Q963T1 ID Q963T1 PRELIMINARY; PRT: 1078 AA.
AC Q963T1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Glutamate-rich protein (Fragment).

```

Search completed: May 13, 2003, 13:56:23
Job time : 44.3021 secs

QY 65 DRYHYNGKVPYDALFSELLMKDPYOLKADIVNEVKGYIIVKDGKYYVYLKDAHA 124
DB 61 DRYHYNGKVPYDALFSELLMKDPYOLKADIVNEVKGYIIVKDGKYYVYLKDAHA 120
QY 125 DNVRTKDEINROKQEHVDNEKVNNAVARSQGRYTTNDGYVFNPADEIETGNAYIVP 184
DB 121 DNVRTKDEINROKQEHVDNEKVNNAVARSQGRYTTNDGYVFNPADEIETGNAYIVP 180
QY 185 HGGHYHYTPKSDLSASELAALAAKAGHAGNMOPSQLSYSS----- 223
DB 181 HGGHYHYTPKSDLSASELAALAAKAGHAGNMOPSQLSYSS----- 240
QY 224 ----- 223
DB 241 SENLQSLKELYDSPAQRYSSEGLVDFPAKIIRTPNGVAIPHGDHYHIFPYKLSAL 300
QY 224 ----- 223
DB 301 EKIARMPVISTGTSTVNTAKPNEVSSLSLSPNSPSSLTSTKELSSASDGYIFNPXDI 360
QY 224 -----TPSPSLPINPGTSHEKHEEDGYGF 247
DB 361 VEETATYIVRHGDHFHYPKSNQIQPTLPNNSLATPSPSLPINPGTSHEKHEEDGYGF 420
QY 248 DANRIIADESGFVMSHGDHNYFFKK 274
DB 421 DANRIIADESGFVMSHGDHNYFFKK 447

RESULT 2

US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-66

Query Match 22.3%; Score 980; DB 4; Length 763;
Best Local Similarity 31.0%; Pred. No. 1e-65;
Matches 266; Conservative 123; Mismatches 223; Indels 292; Gaps 39;
QY 1 CAVALLQCHRS-QENKONNRVSYVDGSSQSKSENLPDQVQSKEGTOAEQIVIKITDQGY 59
DB 1 CSVELGHQAGQVKESNRVSYIDGQAGQKAENLPDDEVSKREGINAECXIKITDQGY 60
QY 60 VTSRGDHYHYNGKVPYDALFSELLMKDPYOLKADIVNEVKGYIIVKDGKYYVYLK 119
DB 61 VTSRGDHYHYNGKVPYDALFSELLMKDPYOLKADIVNEVKGYIIVKDGKYYVYLK 120
QY 120 DAHADNVRTKDEINROKQEHVD-NEKVNNAVARSQGRYTTNDGYVFNPADEIETG 178
DB 121 DAHADNVRTKDEINROKQEHVD-NEKVNNAVARSQGRYTTNDGYVFNPADEIETG 180
QY 179 NAVTVPHGGHYHYIPKSDLSASELAALAAKAGHAGNMOPSQLSYSS-----SL 229
DB 181 DAVTVPHGGHYHYIPKSDLSASELAALAAKAGHAGNMOPSQLSYSS-----SL 240
QY 230 PINPGTSH-----EKH-BEDGYGFDANRIIADESGFVMSHGDEN 268
DB 241 TVTP-TVHQOGENISSLLRELYAKPLSERHVESDGLIEDPAQITRTARGVAVPHGNY 299
QY 269 HYFFKDLTEEQIKAAQKHEEV-----KTSNGLDLSLSHEDYPSNAKEMKDLCKTE 323
DB 300 HF-----IPYEQMSELEKRIARIIPLYRSNHWVPD--SRPEQSPQSTPEPSPQPA- 351
QY 324 EKTAGIMKQYGVKRESIVNKEKNAIYPHGDHHDADPIDEHKPVGIGHSNHYELFKPE 383
DB 352 -----PNPQAPSNPIDE----- 364
QY 384 EGYAKKEGKVVYTGEELTNVVNLKKNSTFNNQNTFLANGOKRVSPFPPELEKKGICINML 443
DB 365 -----KLVK-----EAVRVKG----- 375
QY 444 VKLITPDGKYLEKVGKVFGEVGNIANFELDQYLPQGTFKYTIASKDYPEVSYDGTIT 503
DB 376 -----DGYVFEE-----NGVSR-----YTPAKDLSAETAA-----GIDSKLA 407
QY 504 VPTSLAYKMASQIFIPFHAGDTYLRVNPQFAYPKGTDALVRVDFEHGNAYLENNKVV- 562
DB 408 QESLSHKLGAKKTDLPSSDREFY-----NKAYDILLARHQDLLNKGQVDFEAL 458
QY 563 -----GEIKL-----PI-----PKLNQTTTETAGNKIPVTFMANAYL- 594
DB 459 DNLLERLKVXSKVKLVXDILAFAPIRHPERLGRFNAQITYT-DDEIQVAKLAKYTT 517
QY 595 -----DNQSTYIVVEPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEERK 641
DB 518 EDGYIFDPRDITSDGDAYV--TPHMTSHWIKKDS-LSEAEAAAQAYAK-EKGLTPPS 573
QY 642 TSEKVEKEKLSGTSNSTLEEVPTDVPQVKAKFAESYGMKLENVLFNMDGTIE-- 699
DB 574 TDHQ-----DSGNT-----EAKGAEAIYNRV-KAAKK--VPLDRMPYNLQIYVEVK 616
QY 700 -----LY-LPSEVTK-----KNMADTGEAPQ-----GNG 724
DB 617 NGLIIPHYDHYHNKPEWFEDEGLYEAPKGYTLEDLIATVYVVEHPNRPNSDNGFNA 676
QY 725 EN--KPSENGKVSPTGVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGNVSDP 782
DB 677 SDHVORKNQADNTQTE-KPSEKPKQTERPEBETPREKQSE-----KPESPKPTEE 729
QY 783 MLDSALE-EAPAV--DPVQEKLEK 803
DB 730 PEESPESEEPQVETEKVEKLE 753

RESULT 3

US-08-961-083-56
; Sequence 56, Application US/08961083
; Patent No. 6159469

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/790,912
APPLICATION NUMBER: US/08/790,912
FILING DATE: 29-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,939
FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-401
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1964 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-3

Query Match 4.4%; Score 194.5; DB 2: Length 1964;
Best Local Similarity 22.0%; Pred. No. 9.5e-06;
Matches 165; Conservative 85; Mismatches 220; Indels 281; Gaps 44;

QY 186 GGHYHYIPKSDLSASELAALAHLAGKNQPSQSLY-----SSTP 225
DB 47 GYHYKYVADSELSSEE-----KKQLYDIPTYVENDDETYLVYKLNQSQ 92
QY 226 SPSLPINPOTSKEHE-EDGYCFDANRIIABDESQFVMSHGDHNNHYFFKDLTEQIAAA 284
DB 93 LAELP-NTGSKNERCALVAGASLAALGILI-----FAVS-----KKVKNKTV--- 134
QY 285 QKHLEVKTSNG-----LDLSLSH-----EQDPSNAKEMKDLKKIE-EKAGIMQYQ 334
DB 135 -LHLVLVAGMGVLSVSHALENHLNLYDYLTSGEKLPKREISGYTYIGYREGK 193
QY 335 VKRESIVNKKNAI-----IYPH-GDH-----HHADPIDHEKPPVGIG----- 371
DB 194 TTSDREYSNQESAATPTKQKQVOYNVTNPNFVDHSTVQAICEQTPVSTKPTVQVVEK 253
QY 372 -----HSHSNVELEKPEGVAKREGNKVYT-----GELTNVV 404
DB 254 PFSTELINPRKEEKQSSDSQEQALAEHKNLET-KKEEKISPREKTCVNTLPQDEVLSQ 312
QY 405 N-----LLKNSTN-----NQFTLANGOKRVSEFPPPELEKLGIM-LVKLIT-- 448
DB 313 NKPELLYREETIETKIDFQEEIQENPDIAEGTVR-----KQEGKLOKKVEIVRFSVN 366
QY 449 -----PDGKVLKVSCK--VFGE-----GV-----GNIANFELDQPYLP 480
DB 367 KEVSREIVSTSTTAPSPRIVEKTKTKTKVKEQPETGVEHKDQVSGAIVEPAI-OPELP 425
QY 481 GQTFKYTIASKDYPEVSYDGTVTPTSLAYKMASQTIFFPHAGDTYLRVNPQ-----FAV 536
DB 426 -----EAVSDKGEPEVQP-----TLPEAV-----TDKGET--EVQDESPTVVS 464
QY 537 PKGTDALVAVDFEFHGNAYLENNYKVGEEK--LPIPKL-NQGTTRTAGNKIPVTFMANAY 593
DB 465 DKGEPEQVAPLPYKGN-----LEQVKPETPVKTKKEQGEPT--EEVPV----- 507
QY 594 LDNQSYTYIVEVPILEKENOTDPSILPQKRN---KAQNSKLDKVEPKTSEKV----- 646
DB 508 -----KPTETPVNPNNEGTEGTSI--QEAENPVQFAEESTTNSEKVSPTSSNTGVS 560
QY 647 -----EKEKLSSETGNTSNSTNLEEVPTVDPVQEKVAKFAESYGMKLENVLFN 693

QY 175 EDTGNAYIVPHGGHYHYIPK-SDLSASELAALAHLAGKNQPSQSLYSTSPSPSLPIMP 233
DB 19 EISGYTYI-----GYIREGKTTSESEVSNQSSVATPTKQ-QKVDYNTV--PNFVDHP 68
QY 234 GTSHERHEEDGYFDANRIIABDESQFVMSHGDHNNHYFFKDLTEQIAAAQKHLEVK 293
DB 69 ST-----VOAIEQTPVSESTKTEVQVVEKFEFSTELINPR-----KEKQ 108
QY 294 SHGLDLSLSHEQDPSNAKEMKDLKKIEEKIAGIMQYGVKRESIVYNKKNALIIYPH 353
DB 109 SSDSQEQALAEH-----KNLETKKEEKS-----PKEKTGV----- 138
QY 354 GDHHDADPDE-----HKPVGICHSHSNYELFKPEGVAKREGNKVYTGELTNVNL 407
DB 139 ---NTLNPQDEVLSQGNKPB-----ELLYREETME--TKIDFQEEI----- 174
QY 408 KNSTFNQNTFLANGOKRVSEFPPPELEKLGIM-LVKLIT-----P 449
DB 175 -----QENPDIAEGTVR-----KQEGKLGKKEIVRIFSVMKEVSREIVSTSTTAP 222
QY 450 DGKVLKVSCK--VFGE-----GV-----GNIANFELDQPYLPQGTFFKTTASKDYPEV 496
DB 223 SPRIVERGTRKTQVIREQPETGVEHKDQVSGAIVEPAI-OPELP-----EAVVSDKGEPEV 277
QY 497 SYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQ-----FAVPKGTDALVRVDFEFG 552
DB 278 QP-----TLPEAV-----TDKGET--EVQDESPTVVSQKGEPEQVAPLPYK 320
QY 553 NAYLENNYKVGEEK--LPIPKL-NQGTTRTAGNKIPVTFMANAYLDNQSYTYIVEVPILEK 609
DB 321 N-----IEQVKPETPVKTKKEQGEPT--EEVPV-----KPTETPVNPN 358
QY 610 ENOTDKPSILPQKRN---KAQNSKLDKVEPKTSEK-----VEK 648
DB 359 ECTTEGTSI--QEAENPVQFAEESTTNSEKV-SPTDSSKNTGEVSNPSSDTSVGSN 415
QY 649 EKLSETGNTSNSTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNDGTTIELYLPGEVI 708
DB 416 PEHNSKNSNSEKTVEVP-VNP-----NEGIVE----- 443
QY 709 KKNMADFTGAPQNGENKPSNGKV---STGTVENOPTENKP--ADSLPEAFNKPVPK 763
DB 444 ---GTSNOKETKPVQAPETQTSNGKIANENTGEVSNKPSDKPPVEESNQPERNGTATKP 501
QY 764 ENS-----TDNGMLNPE---GNVGDPMDSALAEAPAVDPVQEKLEKFTASYGLGDSVI 816
DB 502 ENSGNTTSSENQTEPEPSNGNSTEDVSTESNTSNGNEEIKOENE-----LDPDKKV 554
QY 817 FNDGTTIELR 826
DB 555 EEPEKTTIELR 564

RESULT 5
US-08-790-912-3
Sequence 3, Application US/08790912
Patent No. 5976542
GENERAL INFORMATION:
APPLICANT: Weiser, Jeffrey N.
APPLICANT: Plaut, Andrew G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: 1601 Market Street, 36th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

Db 561 SNPSDSTSVGESKNPEHNSKENSEKTYVEVP-VNP----- 597
QY 694 MDGTIELPLSGEVIKKNMADFTGEAPGNGENKPSNGKV---STGTVENQPTENKPP-- 748
Db 598 NEGIVE-----GTSNQETEKVPQPAEETQTSCKIANENTGEVSNKPSDSKPPV 646
QY 749 ADSLPEAPNEKPVKPSN-----TONGMLNPE 775
Db 647 BESNOPEANGTATKPSNCSNTTSENSQTEPE 677
RESULT 6
US-08-790-912-2
; Sequence 2, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZ JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; City: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-790-912-2
Query Match 4.4%; Score 194.5; DB 2; Length 2052;
Best Local Similarity 22.0%; Pred. No. 1e-05;
Matches 165; Conservative 85; Mismatches 220; Indels 281; Gaps 44;
QY 186 GGHYHYPKSDLSASELAARAAHLAGKNWQPSQLSY-----SSTP 225
Db 120 GVHYKYVADSELSE-----KKQLVDIPTVYVNDDETYLYVYKLNQNG 165
QY 226 SPSLPINDGTSHEKHE-EDGYCFDANRIIADESGFVNSHGCHNYFFKDLTTEQKAA 284
Db 166 LAELP-NTGSKNERQALVAGASLAALGILI-----FAVS-----KKVKNKTV--- 207
QY 285 OKHLEEVKTSNG---LDSLSH-----EODYPSNAKEMKDLKKE-ERKIAMQYQ 334
Db 208 -LHLVLVAGMNGVLVSVALENLLNWTYDTELTSGEKPLPKESGYIYIGYREGK 266
QY 335 VKRESIVYNKEKNAT-----IYPH-GDH-----HHADPIDERKPVGIG----- 371

Db 267 TTSDFEVSNOEKSAATPTTKOOKKYVDYNTVTFVDPHSTVQAIQEQTSSTTKPTEVQVVEK 326
QY 372 -----HSHSNYELFPKEGCVAKKEGNKYVT-----GEELTNV 404
Db 327 PFSTELINPRKEEKQSSDSOQLAEHKNLET-KKEEKISFKEKTVNTLNPODEVLSGQL 385
QY 405 N-----LLKNSTFN-----NQNTLANGOKRVSEFPPELEKKLGINM-LVKKLIT-- 448
Db 386 NKPELLYREETIETKIDFQEEIQENPDLAEGTVRV-----KQEGKLGKVKVETVRFISVN 439
QY 449 -----PDGKVLKVSQK--VFGE-----GV-----GNIANFELDOPILP 480
Db 440 KEVSRREIVSTSTAPSRIVEKGTKKTVIKQEPETGVHEKDVQSGAIVERPAI-QPELP 498
QY 481 GQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQIFYPFHAGDYLRVNPQ-----FAV 536
Db 499 -----EAVVSDKGPEVQP-----TLPEAVV-----TDKGET--EVQPESPOTVVS 537
QY 537 PKGTDALVRVDFEHGNAYLENNYKVGEEK--LPIPKL-NQGTTRTAGNKIPYTFEMANAY 593
Db 538 DKGEPEQVAPLPEYKGN-----TEQVAPETPVKTRKQGPET--EEVPV----- 580
QY 594 LDNQSTYIVIVEPILEKENQTDKPSILPQFKRN---RAQENSKILDERVEEPKTSKV--- 646
Db 581 -----KPTETPVNPEGTTEGTSI--QEAENPVQPAEESTTNSKVSPTDTSSENTGEVS 633
QY 647 -----EKEKLSGTGNTSNTSILEEYVTVDPVQEKVAKFAESYGMKLENVLEN 693
Db 634 SNPSDSTSVGESKNPEHNSKENSEKTYVEVP-VNP----- 670
QY 694 MDGTIELPLSGEVIKKNMADFTGEAPGNGENKPSNGKV---STGTVENQPTENKPP-- 748
Db 671 NEGIVE-----GTSNQETEKVPQPAEETQTSCKIANENTGEVSNKPSDSKPPV 719
QY 749 ADSLPEAPNEKPVKPSN-----TDNGMLNPE 775
Db 720 EESNQPEKNGTATKPSNCSNTTSENSQTEPE 750
RESULT 7
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 4.1%; Score 181.5; DB 4; Length 10182;
Best Local Similarity 19.0%; Pred. No. 0.00all;
Matches 177; Conservative 147; Mismatches 391; Indels 219; Gaps 42;
QY 6 NQHRSOENKDNRRV---SYVDGSSQSSEKSENTPDQVSO-----KEGIAQEVIVIKI 54
Db 5345 NEDSTOKTVDNAIDNGSTVITQCHNPENLKNSTIDQTSIRINTAKNDLHGVEKLQ----- 5399
QY 55 TDGYVTSHGHCHHYHNGKVPYDALFSEELLKDPNQYKADIVNEVG-GYIIKVDOK 113
Db 5400 RDRGTANQELGQLGYLND--PKS--GEESLVNGSNTSRSEVEHLNKAESLNNAMK----- 5451

QY 114 YVYLKDAHAAHNVRTKDEINRQKQEHVKD-NEKVNNSVAVARSQGRYTTINDGVVENPA 171
Db 5452 ---QLRDKVAERTNVKQSSDYNDSTEHORGVDQALQAEANIINEIGNPTLNKSEIEQKL 5508
QY 172 DIIEDTGNAYIVPHGGHYIYIPKSDI-----SASELAAKAHAKHAKNQPSOLSYS----- 222
Db 5509 QQLTDAQNAL---QGSHLLEEAKNNAITGINKLTALNDAQRKAIENVCQAOTIPAVANQ 5565
QY 223 ---STPSPLNPOTSEKHEEDGY-----GFDANRIAEDESFGVM 262
Db 5566 LTLDRINTAMQALRDKVQCNVNIHQSNYFNEDEPKPHNDNSVQAGCTIIDKLODFIM 5625
QY 263 SHGDHNYFFKDLBEEQKAAQK-HLEEVKTSH---NGLDLSLS-----SHEQDYPNAKEM 315
Db 5626 NKNEIEQAINQINTTQTALSGENKLTDOESTNRQTEGLSSLNTAQINAQKDLVNAQKTR 5685
QY 316 KOLDKKI---BEKIA-----GIMKQGVKRESIVVAKENKAIYIPHGDHHPIDEH 365
Db 5686 TDAQAALAAKEINSAMNLRGIONKEDIKRSSAYINADPTKYTAYDQALQNAENIINA 5745
QY 366 KPVGIGHSHSNYELFKPEEGVAKKQKQYITGEELTNVNVNLLKNS-----TFNNQNFLL 419
Db 5746 TP-----NVELNK---ATTIEQALSRVQQAQQLDGVQOLANAKQAQTIVNGLN-SL 5793
QY 420 ANGOKRVSPFPELEKKGILNKLITPDGKYLEKVS-----KVFGEVGNIAN 471
Db 5794 NDGOKR-----ELNLLINSANTRTKVQEEELNKATELNHAMEALRNSVONVDQ 5840
QY 472 FELDQPLPQGTFFKTIASKYPEV-SYDGTFTVPTSLAYKMASOTIYFPFHAGTYLVR 530
Db 5841 VKOSSNV-----NEDQPEQHNVDN-----AVNEAQATI-----NNAQPV 5876
QY 531 NPQFAPKGTDALVRVDFEFGHNAVLENNYKVEIKL-----PIPKLNQGT 577
Db 5877 LDKLATERITQVNTTKDALHAQAKLTODQAAETGIRGLTSLNEPQKNAEYAKVTAAT 5936
QY 578 RTAGNKI---PVTFMANAYL-----DNQSTIYVE-----VPILE 608
Db 5937 ROBRVNIROEATLTDLMLGLRSIKDKNDKNSKYINEDHDQOQAYDNVNAQVVID 5996
QY 609 KENQDKPSILPOFRANKAQENSKL-DEKVEEPTSEK---VEKEKSETGNSTNSL 663
Db 5997 ETQATLSSSTINQLANAVTAQAKNSLHGDTKLOHDKDSAKQTAQLQNLNSAKKHMEDSLI 6056
QY 664 EEPVTPDQVEKVAFAESYGMKLENVLFNMDGTIELY---LPSGEVI-----KKNMADF 715
Db 6057 DNESTRTQVQHLTE-----AQLDGLMGALAKESIKDYNIVSNGNYINAPSKKQAYDA 6111
QY 716 TGEAPQ--NGENKPSN-GKYSTG--TVENOPTENKPADSLPEAPN--EKPVKPENSTD 768
Db 6112 AVQNAQNIINGTNOPTINKGNVTTATQVTKNTKDALDGDHRLLEAKNANQTIIRLSNLN 6171
QY 769 NGMLNPEGVNGSDPMLDSALEAPAVDPVOEKLE 802
Db 6172 NACKDREKNL-----VNSASTLEQVOQNLQ 6196

RESULT 8

US-08-242-932-2
Sequence 2, Application US/08242932
Patent No. 5595740
GENERAL INFORMATION:
APPLICANT: Brady, L. Jeannine
TITLE OF INVENTION: Cloning of No. 5595740-Iga Fc Binding Forms of
the Group B Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA

ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/242.932
FILING DATE: 16-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF142
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 984 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
us-08-242-932-2

Query Match 4.1%; Score 180.5; DB 1; Length 984;

Best Local Similarity 18.1%; Pred. No. 3.9e-05;
Matches 13; Conservative 109; Mismatches 246; Indels 255; Gaps 32;

QY 194 KSDLSASELAATAKHAHLA-----GKNMQPSQLSYSSPSPSLPI-----NPG 234
Db 5 KDSVKTTEVAAKPYSPMAQTDOGNSSSSELETKMEIPTDIKRAVEVKTAGTISA 64
QY 235 TSHKHEHEDGYGDANILIAEDSGFVMSHGDNHFFKDLTEEQIKAKQKLEEVKTS 294
Db 65 TDTCKREKQLOQKNN--LKNDVDNTILSHEQKNE--FKTKIDETNDSALLENOFNE 120
QY 295 HNGLDLSSSHQ-DYPSNAKEMKDL---DKKIEKIAGIMKQYGVKRESI-----VV 342
Db 121 TNRLHLIKQHEEVEKDKAKQKQTKQSDTKVD--LSNIDKELNHOKSQVEKMAEQGIT 178
QY 343 NKENAIYIPHG--HHADPIDEHKPVGI-----GHSNHNELFKPEEGVAKKQKRVYT 396
Db 179 NEDKDSMLKIEDIRKQAQQAQKDEAEVAVQLEEEAHSKL-----KQVVEDFRKKFKT 232
QY 397 GEELTNVNLKNSFTENQNTFLANGOKRVSFSPFPPELEKKGILNMLVKLITPDGKVLK 456
Db 233 SEQVTPKRVKRDLAANE-----NNQKIETVSP-----253
QY 457 VSGKVFGEVGNIANFELDOPYLPQGTFFKTIASKQYPEVSYDGTFTVPTSLAYKMASQT 516
Db 264 -----NITVE-----GEDVKFTVAKSDSKTLD-----288
QY 517 IFYFPFHAGTYLVRNPOFAVPKGTDALVRVDFEFGHNAVLENNYKVEIKLPIPKLNQGT 576
Db 289 -----FSDLLTKYNPSV-----DRISTNYKNT---DNHKAETIKLKLKLNESQ 331
QY 577 TRT-----AGNKIPVTFMANAYLDNOSTYIIVEVPILEKENQTDKPSILPQKRNKAQEN 630
Db 332 TVTLKAKDSDSGNVVEKTF-----TITVQKKEK-----QVEKTPQEKD 369
QY 631 SKLDEKV-EEPKTSKVEKEKLSGTGN-----STNSSTILEEVPTV-----659
Db 370 SKTEEYQPEPKSNDKNOLOELIKSAQOELEKLEKALKELMEOPEIPSPNPEYGIQSIWE 429
QY 670 ---DPVQEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSEVGIKKNMAD 714
Db 430 SQKEPIOEATISFKKIIGDSSSKYTYTEHYENKYKSDPMYQLHAQM-----EMLTKVVQ 484
QY 715 FTGAPQGNENKPSKNGKST-----GTVENOPT-----NK-----PADSL 752
Db 485 YMKYPDPAEIKKIFESDMRKREDNYGSLNDALKGYPEKFTLTPFNKIKQIVDLDKK 544

QY 753 PEAPNEKPVKPNSTONGMLNPGNVGSDPMLDSALEEA-----PAVDPVQEKLEK 803
Db 545 VEOQDPAPI-PENSE-----MDQAKERAKIAVSKYMSKVLGVDGVHQHLOK 587
QY 804 FTASYGLGL-----DSVIFNMD 820
Db 588 KNSKIVDLFKLEAIKQOTIFDID 612

RESULT 9
US-08-714-481-2
; Sequence 2, Application US/08714481
; Patent No. 576606
; GENERAL INFORMATION:
; APPLICANT: Brady, L. Jeannine
; TITLE OF INVENTION: Cloning of No. 576606-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/714,481
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-714-481-2

Query Match 4.1%; Score 180.5; DB 1; Length 984;
Best Local Similarity 18.1%; Pred. No. 3.9e-05;
Matches 135; Conservative 109; Mismatches 246; Indels 255; Gaps 32;

QY 194 KSDLASELAARAAHIA-----GKNMQPSQLSYSTPSPSLPI-----NPG 234
Db 5 KDSVKITTEVAAPKPYPSMAQTDOGNSSSELETTIKNEIPTDIDKKAPEVKEKTAGETSA 64
QY 235 TSBEKEHEEDGYGDANRIIAEDSGFVMSHGDHNYFFKKDLTEEQIKAAQKHLEEVKTS 294
Db 65 TDGKREKQLQOQKNN--LKNDVDNYNILLSHEQKNE--FKTKIDETNDSALLELENOFNE 120
QY 295 HNGLDLSLSHEQ--DYSNAKEMKDL--DKKIEEKIAGIMQGVKVRRESI-----VV 342
Db 121 TNRLLHIKQHEEVEKDKKAKQOQRTKQSDTKVD--LSNIDKELNKHQKSOVKEMAEQKGIT 178
QY 343 NKEKNALIIYPHG--HHADPIDRHKPVGI-----GHSNHYELFKPEEGVAKKGNKVT 396
Db 179 NEKDSMLKIEIDIRQAQQAQKEDAEVQVQLEEFASHL-----KQWEDFRKKFT 232
QY 397 GEELTNVNLKSTNNQNTFLANGQKRVSFSPFPELEKKGILNMLVKLITPDGKVLK 456

Db 233 SEQVTPKRVKRDLAANE-----NQOKIELTVSPE----- 263
QY 457 VSGKVEGEGVGNIAFELDQPYLPQOTFKYITIASDYDEVSVDGFTFTVPTSLAYKMASQT 516
Db 264 -----NITVYE-----GEDVKFTVAKSDSKTILD----- 288
QY 517 IFYPHAGDTYLRVNPQFAVPAGTDAALVRVDFDEHGNAYLENNYKVGKIKLPKPLNGOT 576
Db 289 -----FSDLLTKYNPSVS-----DRISTNYKNTN---DNHKAIEITIKNLKLESQ 331
QY 577 TRT-----AGNKIPVTIFMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOFKRNKAGEN 630
Db 332 TVTLKAKDDSGNVEKTF-----TIIVQKKEK-----QVPKTIPEQKD 369
QY 631 SKLDEKV--EETKTSKVEKEKLSSETGN-----STNSNLTLEEVTV----- 669
Db 370 SKTEEKVPQEPKSNKNOLOELIKSAQOELEKLEKAIKELMEQPEIPSNPEYGIQKSIWE 429
QY 670 ---DVOQEKVAKF-----AESYGMKLENVLFNMDGTIELYLPSEGVKKNHAD 714
Db 430 SOKEPIQIAITSFKKLIIGDSSSKYITEHYFNKYKSDFNQYQLHAOM-----EMLTRKVVQ 484
QY 715 FTGEAPQNGENKPSENGKSVT-----GTVENQPTPE-----NK-----PADSL 752
Db 485 YMNKYPDRAETIKKIFESDMKRTKEDNYGSLNDALKGYFEKYFLTPTFNKIKOIVDDLQK 544
QY 753 PEAPNEKPVKPNSTONGMLNPGNVGSDPMLDSALEEA-----PAVDPVQEKLEK 803
Db 545 VEOQDPAPI-PENSE-----MDQAKERAKIAVSKYMSKVLGVDGVHQHLOK 587
QY 804 FTASYGLGL-----DSVIFNMD 820
Db 588 KNSKIVDLFKLEAIKQOTIFDID 612

RESULT 10
PCT-US95-06111-2
; Sequence 2, Application PC/TUS9506111
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Street address: 186 Grinter Hall
; APPLICANT: City: Gainesville Florida
; APPLICANT: State/province:
; APPLICANT: Country: US
; APPLICANT: Postal code/Zip: 32611
; APPLICANT: Phone number: 904-392-8929
; APPLICANT: Fax number: 904-392-6600
; APPLICANT: Telex number:
; TITLE OF INVENTION: Cloning of Non-IgA Fc Binding Forms of
; TITLE OF INVENTION: the Group B Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: PCT/US95/06111
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/242,932
; FILING DATE: 16-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.

```

; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UFI142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-06111-2

Query Match 4.1%; Score 180.5; DB 5; Length 984;
Best Local Similarity 18.1%; Pred. No. 3.9e-05;
Matches 135; Conservative 109; Mismatches 246; Indels 255; Gaps 32;

Qy 194 KDLASASALAAKHAHLA-----GKNQPSQLSYSTPSPSLPI-----NPG 234
Db 5 KDSVKITEVAAPYPSMAQTDOGNSSSELETTKMEIPTTDIKKAVEPVEKTAGTSA 64
Qy 235 TSHEKHEEDGYGDARIIAEDSGVMSGHGHNHYFFKKDLTEQIKAAQKHLEEVKTS 294
Db 65 TDTCKREKQLQOWKNN--LKNDVDNITLSHEOKNE--FKTKIDETNDSALLELENOFNE 120
Qy 295 HNGLDLSLSHEO--DYPSNAKEMKDL---DKKLEEKIAGIMKQGVYKRESI-----VV 342
Db 121 TNRLLHIKHEEVEKKAKKQOATLKOSDTKYD--LSNIDKELNHOKSOVKEMAEOKGIT 178
Qy 343 NKEKNALIIYPHGD--HHADPIDHKKPVGI-----GHSNHYELFKPEEGVAKKEGKVTY 396
Db 179 NEDKOSMLKIEDIRKQAOQADKEDAEVKVQLEESAHSKL-----KQWVEDFKKKT 232
Qy 397 GBLTNVNVLLKNSFNQNTLANGOKRVSFSPPELEKGLINMLVLIPTDQGVLEK 456
Db 233 SEQVTPKRVKRLAANE-----NNQOKTELTVSPE-----263
Qy 457 VSGRVFEGVGNIAFPDQPLYPGQTFKVTIASKDYPEVSYDGTFTVPTSLAYKMASQT 516
Db 264 -----NITVE-----GEDVRFVAKSDSKTILD-----288
Qy 517 IFYFPHAGDYLRVNPQAFVKGTDALVRVDFEFGNAYLENNYKVEIKLPIPKLNQGT 576
Db 289 -----FSDLLTKYNPSVS-----DRISTNYKNT--DNHKAETITIKNLKLSQ 331
Qy 577 TRT-----AGNKIPVTFMANVLDNQSTVIVBVPITLKENQTDKPSILPQFRRKKAQEN 630
Db 332 TVTLKAKDDSGNVVEKTF-----TITVQKKEK-----QVPTPEQKD 369
Qy 631 SKLDEKY--EEPKTSEKVEKELSETGN-----STNSNLEEVPV-----669
Db 370 SKTEKVPQEPKSNKQLOELIKSAQOOLEKLEKATKELMEQPEIPSNPEYGIQKSIWE 429
Qy 670 --DPQVEKAKP-----AESYCKLENVLFNMGGTIELYLPSPGVIKKNWAD 714
Db 430 SQKEPIQEAITSFKKIIGDSSSSYYTHYENKYSKPFMPNVLHAQM-----EMLTKEVVO 484
Qy 715 FTGAPOGNGENKPSNGKYST-----GIVENQTE-----NK-----PADSL 752
Db 485 YMKYVPDAEIKKIFESDMKRTKEDNYGSLNDALKGFEKYEKLTLPENKIKQIVDLDKK 544
Qy 753 PEAPNEKVPKPNSTDNGLNPNCGVSDPWLDSALEEA-----PAVDPOVEKLEK 803
Db 545 VEQDPAPPI--PENSE-----MDQAREKAKIAVSKYMSKVLQGVQHOHLQK 587
Qy 804 FTASYGLGL-----DSVIFNMD 820
Db 588 KNSKIVDLKLELAIAKQIIFDID 612

RESULT 11
us-08-296-791-6
; Sequence 6, Application us/08296791
```

```

; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; US-08-296-791-6

Query Match 4.1%; Score 180.5; DB 4; Length 1848;
Best Local Similarity 17.9%; Pred. No. 9.9e-05;
Matches 165; Conservative 128; Mismatches 345; Indels 283; Gaps 35;

Qy 12 ENKNNRYSYVDGSG--SSQKSENLTPOVSOKEGICABQIIVIKITDQGYVTSHGDIHYIY 70
Db 470 KCKNEGLKVGDTGTVLKOKADANNKVAQFSQVGLVSGRSTLVLLNDKQVDPNSIYFGFR 529
Qy 71 NGKVPEYDA-----LFSEELMKDPN-----YOLKQADI 98
Db 530 GORLDLNGSLTFDHIRNTDDGARVVNHNMTNTSNTITGESLITNPNTITSYNIEAQQD 589
Qy 99 VNEV-----KGGHIIKVDGKYVYVVKDAHADNV-----RTKDEINRQ 136
Db 590 DHPLEIRSIPIYQLFENQDNKRSYTLTKKGASTRSELPQNSGESNENWLYMGTSDAAKRN 649
Qy 137 KCEHVKNDEKVASNVAVARSQGRVYTNDCYVFNPAIIIEDTGNNAVIVPHGGHYHYIPKSD 196
Db 650 VVNHI--NERNNG-----FNGYGEETKATQNGKLVNTENG-----KSD 688
Qy 197 LSASELAARAKAHLAKGNKQPSQLSYSTPSPSLPNTPGTSHEKHEEDGYGDANRIIAD 256
Db 689 QNRFLITGTNLNGDLNVEKGTFLSGRPTPHARDIAGISSTK--KDPHFTENNEVVVED 746
Qy 257 E-----SGFVMSGHGDHNYFFKKDLTEQIKAAQKH 287
Db 747 DMINRFKATTVNTGNASLYSGRVNANITSNITASNNAQVHIGYK---TGDTCVRSYD 803
Qy 288 LEEKVTSHNGLDLSLSHEEDYPSNAKEMKDLKKI-----EKTAGIMKQGVKRESIVV 342
Db 804 TGVV--TCHNSNLSEXALNSFNPTNLRGVNVNTENASFTLGRANLESTIGISIGTSQ----V 858
Qy 343 NKEKNALIIYPHGDHHAADPIDHKKPVGI--GHSNHYELFKPEEGVAKKEGKVTYGEEL 400
Db 859 NLKENS-----HWHLTGNSNVNQLNLTNGHILN-----AONDANKVITYNTL 901
```


QY 401 TNNVLLKNSTF-----NNONFTLANGQKRVSFSPPELEKKIGINMLVKLITPDGK 452
Db 902 T-VNSLSGNGSYVYVDFTNKSNKVVN--KSATGNETLQVADKTGFENHNELTLFDAS 958
QY 453 VLEKVGKRVGSGVGNIAFELDQYPLGQTFKYTIASKD-----YPEVSYDGTFTVPT 506
Db 959 NATRNNEV-----TLANGSVDR-----GAWKYKLRVNGRYDLYNPEV-----997
QY 507 SLAYKMASOTIFYPFHAGDTYLRVP--QFAVPKGTDALVRVDFEFGHNAVLENNYKVG 563
Db 998 ----EKRNTV-----DTNITTPNDIQADAPSAQ-----SNNEEIA 1030
QY 564 EIKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQSTYVEVPILEKENOTDKPSILPOFK 623
Db 1031 RVETVPPAPATSAIASEQPET-----RPAETAQPA-----1063
QY 624 RNKAQENSKLDEKVEEPTSEKVEKELSTGNSTLSNTELEEVPTDVPQEKVAKFAESY 683
Db 1064 -----MEETNTANSTETAPKSDTATQNTENPNSVPS-----ETTEKVAENP 1105
QY 684 GMLKLVFNMDGTIELYLPSEGVTKKNMADF-----TGEAPQNGENKPSGKSVSTGT 738
Db 1106 POENETVAKNEQATEPTPQNGEVAKEDQPTVEANTQTNEATQSGKTEETQ-----TAE 1160
QY 739 VENOPT-----ENKPADSLPEAPNEKPV-----KPNSTONGMLNPGNVGSDPM-----783
Db 1161 TKSEPTESVTVSENQPEKTVSQSTEDKVVVEKEKAKVETEETQKAPQVTSKEPPKQAP 1220
QY 784 -----LDSALEEAPAVDPVQ 798
Db 1221 APEEVPDTNAAEAQALQQTQ 1241

RESULT 12

PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Irecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RET
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid

TOPOLOGY: unknown
PCT-US95-10661A-6
Query Match 4.1%; Score 180.5; DB 5; Length 1848;
Best Local Similarity 17.9%; Pred. No. 5.9e-05;
Matches 165; Conservative 128; Mismatches 345; Indels 283; Gaps 35;
QY 12 ENKONNRVSYVDGSG--SSQKSENLTDPQVQKEGIAEQIVIKITDQGYVTSHGHDHYHY 70
Db 470 KGKNEGLLKVDGIVLKQRADANNKQAFSGVIGSRSTLVLNDDQVDPNSIFGFR 529
QY 71 NGKVPYDA-----LFSEELMKDPN-----YOLKDAADI 98
Db 530 GGRDLNCGNSLTFDIRNIDDGARVVNHNMTNITSNITITGESLTNPNTITSYNIEAQDD 589
QY 99 VNEY-----KGGYIILKVDGKYVYLKDAHADNV-----RTKDEINRQ 136
Db 590 DHPRIIRISIPRQLFYQDNRSYITLKGASTRSELSPQNSGESNENWLYWGRTSDAAKRN 649
QY 137 KOEIVKONEKYNVAVARSQGRYTTNDGVYFNPADIIEDTGNAYIVPHGHGHHYIPKSD 196
Db 650 VMNHI--NNERNMG-----FNGYFGEETKATQNGKLNVTFG-----KSD 688
QY 197 LSASELAAKAHLAGKKNOPSQLSYSTPSPSLPINPTGSHEKHEEDGYGFDANRIIAED 256
Db 689 QNRLLTGGTNLGDNLNVEKGTIFLSCRPTPHARDIAGISSTK--KDPHTENNEVVVED 746
QY 257 E-----SGFVNSHGCDHNHYFFRKDLTTEQIKAAQKH 287
Db 747 DWINRNEKATTMNVTGNASLYSGRNVANITSNITASNAQVHGYK---TGDVTCVRSDY 803
QY 288 LEEVKTSHNGLDLSLSSHEQDYPSSNAKEMKOLDKI-----EETIAGIMKQYGVKRESIV 342
Db 804 TGYV--TCHNSNLSEKALNSPNTNLRGNVLNLTENASTLTKANLFGTIQIGTSQ---V 858
QY 343 NKEKNAIYPHGDHHDADPIDEHKPVGI--CHSHSNTELEKPEBEGVAKKCGNKVYTGEL 400
Db 859 NLKENS-----HWLGTGNSNVQLNLTNGHILN-----AQDANKVTVTNTL 901
QY 401 TNNVLLKNSTF-----NNONFTLANGQKRVSFSPPELEKKIGINMLVKLITPDGK 452
Db 902 T-VNSLSGNGSYVYVDFTNKSNKVVN--KSATGNETLQVADKTGFENHNELTLFDAS 958
QY 453 VLEKVGKRVGSGVGNIAFELDQYPLGQTFKYTIASKD-----YPEVSYDGTFTVPT 506
Db 959 NATRNNEV-----TLANGSVDR-----GAWKYKLRVNGRYDLYNPEV-----997
QY 507 SLAYKMASOTIFYPFHAGDTYLRVP--QFAVPKGTDALVRVDFEFGHNAVLENNYKVG 563
Db 998 ----EKRNTV-----DTNITTPNDIQADAPSAQ-----SNNEEIA 1030
QY 564 EIKLPIKLNQGTTRTAGNKIPVTFMANAYLDNQSTYVEVPILEKENOTDKPSILPOFK 623
Db 1031 RVETVPPAPATSAIASEQPET-----RPAETAQPA-----1063
QY 624 RNKAQENSKLDEKVEEPTSEKVEKELSTGNSTLSNTELEEVPTDVPQEKVAKFAESY 683
Db 1064 -----MEETNTANSTETAPKSDTATQNTENPNSVPS-----ETTEKVAENP 1105
QY 684 GMLKLVFNMDGTIELYLPSEGVTKKNMADF-----TGEAPQNGENKPSGKSVSTGT 738
Db 1106 POENETVAKNEQATEPTPQNGEVAKEDQPTVEANTQTNEATQSGKTEETQ-----TAE 1160
QY 739 VENOPT-----ENKPADSLPEAPNEKPV-----KPNSTONGMLNPGNVGSDPM-----783
Db 1161 TKSEPTESVTVSENQPEKTVSQSTEDKVVVEKEKAKVETEETQKAPQVTSKEPPKQAP 1220
QY 784 -----LDSALEEAPAVDPVQ 798
Db 1221 APEEVPDTNAAEAQALQQTQ 1241

RESULT 13

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-4

Query Match 4.1%; Score 180; DB 2; Length 1435;
Best Local Similarity 19.3%; Pred. No. 7.4e-05;
Matches 113; Conservative 95; Mismatches 197; Indels 182; Gaps 26;

QY 11 QENKDNRRVSYVDGSSQSSQSENLTDP-----QVSQKEGIAEQIIV--KITDQGYVTSH 63
Db 789 QKERDDSLSKISVSPENSRPETDAXDTSNLLKLGKGVDSMPKAVIGSSPNDINVTQ 848
QY 64 GDHYHYNGKVPYDALFSEELLKMDPNYQLKADIVNEV----- 102
Db 849 GDNISGVNKPISDDVRPKKELEDQNSDESEETVVNHISKSPSINNGDDSGSATVSE 908
QY 103 -----KGGYIIKVD--GKYVYVYLKDAHADNVRTKDEINRQKQEHVKDNEKVNNAVARS 156
Db 909 SSSNTGLSDDRRNGDTFVRTQDTANTEDVIRKENADKDEKGADEHERHSTSESLSP 968
QY 157 QGRYTTND--GYVFNPAIIEDTGN--YIVPHGGHYHYIPKSDLS-----ASELAAGA 207
Db 969 EEMLTNDEGGNSLNHEEYKEHTSNSDNYQSGGIYVNMVNEKELKDTLENPSSSLDEGKA 1028
QY 208 H--LAGKNMQPSQLSYSTSPSPSLPINCPTSHE----- 238
Db 1029 HEELSEPNLSDDQ--DMSNTPGPL-----DNTSEETTERISNNEYKVNREDEERTLTKEYED 1083
QY 239 -----KHEEDGYGFDANRRIA--EDESGFV-----MSHGDHNYHFF---KK 274
Db 1084 IYLVKSHMNRSDGDELVDNSDLTVNDESDAEAKMKGNDTSEMNSHSSQHIESDQOKN 1143
QY 275 DL-----TEEQIAAQAQ-----HLEEVKTSHNGLDLSLSHEDQY 308
Db 1144 DMKTVGDLGTHVQNEISVPVTGEIDEKLRKESKSIHKAEEELSHTDIHKI--NPEDR 1201
QY 309 PSNAKEMKOLDKKIEKIAKIMQYGVKRESIVVANKENKAIYPRGDHH-----HADPIDE 364
Db 1202 NSNTLHLKDIRNEENER-----HLTNQINISQERD--LQKHGFHTMNLHGDGVSE 1251
QY 365 HRPVGIHSHSNYELFKPEGVAKKGNKVYTGEBELTVNVLNKKNSTNNQNTFLANGQK 424
Db 1252 RSQ--INSH-----HGNRQDRGNSGNVLMRS-----NNNNFN--NIPS 1288
QY 425 RVSFSPPLERKLGIM-----LVKLITPDGKVLKGVSK 460
Db 1289 RYNL-----YDKKLDLDLYENRNDSTTKELIKLAEINKCENEISVK 1330

RESULT 15

US-08-487-826B-4
Sequence 4, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chaitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487.826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPL1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-4

Query Match 4.1%; Score 180; DB 2; Length 1435;
Best Local Similarity 19.3%; Pred. No. 7.4e-05;
Matches 113; Conservative 95; Mismatches 197; Indels 182; Gaps 26;

QY 11 QENKDNRRVSYVDGSSQSSQSENLTDP-----QVSQKEGIAEQIIV--KITDQGYVTSH 63
Db 789 QKERDDSLSKISVSPENSRPETDAXDTSNLLKLGKGVDSMPKAVIGSSPNDINVTQ 848
QY 64 GDHYHYNGKVPYDALFSEELLKMDPNYQLKADIVNEV----- 102
Db 849 GDNISGVNKPISDDVRPKKELEDQNSDESEETVVNHISKSPSINNGDDSGSATVSE 908
QY 103 -----KGGYIIKVD--GKYVYVYLKDAHADNVRTKDEINRQKQEHVKDNEKVNNAVARS 156
Db 909 SSSNTGLSDDRRNGDTFVRTQDTANTEDVIRKENADKDEKGADEHERHSTSESLSP 968
QY 157 QGRYTTND--GYVFNPAIIEDTGN--YIVPHGGHYHYIPKSDLS-----ASELAAGA 207
Db 969 EEMLTNDEGGNSLNHEEYKEHTSNSDNYQSGGIYVNMVNEKELKDTLENPSSSLDEGKA 1028
QY 208 H--LAGKNMQPSQLSYSTSPSPSLPINCPTSHE----- 238
Db 1029 HEELSEPNLSDDQ--DMSNTPGPL-----DNTSEETTERISNNEYKVNREDEERTLTKEYED 1083
QY 239 -----KHEEDGYGFDANRRIA--EDESGFV-----MSHGDHNYHFF---KK 274
Db 1084 IYLVKSHMNRSDGDELVDNSDLTVNDESDAEAKMKGNDTSEMNSHSSQHIESDQOKN 1143
QY 275 DL-----TEEQIAAQAQ-----HLEEVKTSHNGLDLSLSHEDQY 308
Db 1144 DMKTVGDLGTHVQNEISVPVTGEIDEKLRKESKSIHKAEEELSHTDIHKI--NPEDR 1201
QY 309 PSNAKEMKOLDKKIEKIAKIMQYGVKRESIVVANKENKAIYPRGDHH-----HADPIDE 364
Db 1202 NSNTLHLKDIRNEENER-----HLTNQINISQERD--LQKHGFHTMNLHGDGVSE 1251
QY 365 HRPVGIHSHSNYELFKPEGVAKKGNKVYTGEBELTVNVLNKKNSTNNQNTFLANGQK 424
Db 1252 RSQ--INSH-----HGNRQDRGNSGNVLMRS-----NNNNFN--NIPS 1288
QY 425 RVSFSPPLERKLGIM-----LVKLITPDGKVLKGVSK 460
Db 1289 RYNL-----YDKKLDLDLYENRNDSTTKELIKLAEINKCENEISVK 1330

Tue May 13 14:29:42 2003

us-09-471-255-16.ra1

Page 12

Search completed: May 13, 2003, 13:58:10
Job time : 31.0205 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:53:57 ; Search time 20.7625 Seconds
(without alignments)
3723.142 Million cell updates/sec

Title: US-09-471-255-16

Perfect score: 4396

Sequence: 1 CAYALNQHRSQENKDNNRVS.....GTIELRLPSGEVIKKNLLIS 840

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 200000000
Maximum DB seq length: 200000000

.....

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: /cgn2_6/ptodata/2/pubpaa/us08_NEW_PUB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/FCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/FCTUS_PUBCOMB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/us08_NEW_PUB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/us09_NEW_PUB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/us09_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/us10_NEW_PUB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/us10_PUBCOMB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	4396	100.0	840	9	US-09-884-465A-10	Sequence 10, Appl
2	4376.5	97.3	1039	9	US-09-884-465A-6	Sequence 36, Appl
3	3402.5	77.4	1152	9	US-09-884-465A-379	Sequence 379, App
4	3402.5	77.4	1328	9	US-09-884-465A-381	Sequence 381, App
5	3402.5	77.4	1365	9	US-09-884-465A-382	Sequence 382, App
6	3402.5	77.4	1378	9	US-09-884-465A-378	Sequence 378, App
7	3237	73.6	906	9	US-09-884-465A-370	Sequence 370, App
8	3228	73.4	906	9	US-09-884-465A-332	Sequence 332, App
9	3218	73.2	906	9	US-09-884-465A-372	Sequence 372, App
10	3215	73.1	900	9	US-09-884-465A-334	Sequence 334, App
11	3200.5	72.8	901	9	US-09-884-465A-334	Sequence 334, App
12	3193.5	72.6	633	9	US-09-884-465A-350	Sequence 350, App
13	3193.5	72.6	633	9	US-09-884-465A-354	Sequence 354, App
14	3193.5	72.6	633	9	US-09-884-465A-361	Sequence 361, App
15	3193.5	72.6	633	9	US-09-884-465A-362	Sequence 362, App
16	3191	72.6	900	9	US-09-884-465A-352	Sequence 352, App
17	3184.5	72.4	633	9	US-09-884-465A-351	Sequence 351, App
18	3184.5	72.4	633	9	US-09-884-465A-352	Sequence 352, App
19	3184.5	72.4	633	9	US-09-884-465A-357	Sequence 357, App

Sequence 358, App
Sequence 359, App
Sequence 360, App
Sequence 361, App
Sequence 362, App
Sequence 363, App
Sequence 364, App
Sequence 365, App
Sequence 366, App
Sequence 367, App
Sequence 368, App
Sequence 369, App
Sequence 370, App
Sequence 371, App
Sequence 372, App
Sequence 373, App
Sequence 374, App
Sequence 375, App
Sequence 376, App
Sequence 377, App
Sequence 378, App
Sequence 379, App
Sequence 380, App
Sequence 381, App
Sequence 382, App
Sequence 383, App
Sequence 384, App
Sequence 385, App
Sequence 386, App
Sequence 387, App
Sequence 388, App
Sequence 389, App
Sequence 390, App
Sequence 391, App
Sequence 392, App
Sequence 393, App
Sequence 394, App
Sequence 395, App
Sequence 396, App
Sequence 397, App
Sequence 398, App
Sequence 399, App
Sequence 400, App

ALIGNMENTS

```

RESULT 1
US-09-884-465A-10
; Sequence 10, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus AN
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/88
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-10

```

QY	1	CAVALNOHRSQENKDNRRVYVDSQSSQKSENITPDQVQSQKEGIAEQAEIVIKITDQGVV	60
Db	1	CAVALNOHSSQENKDNRRVYVDSQSSQKSENITPDQVQSQKEGIAEQAEIVIKITDQGVV	60
QY	61	TSHGDRHYHYNGKVPYDALFSEELLMKDPNVLQDADIVNEVKGVIKVDGKYVYVLKD	120
Db	61	TSHGDRHYHYNGKVPYDALFSEELLMKDPNVLQDADIVNEVKGVIKVDGKYVYVLKD	120
QY	121	AAHADNVRTKDEINPQKOEHVKDNEKYNVNAVARSOGRYTTNDGVYFNPADIIEDTGA	180
Db	121	AAHADNVRTKDEINPQKOEHVKDNEKYNVNAVARSOGRYTTNDGVYFNPADIIEDTGA	180
QY	181	YIVPGHGHHYIPKSDLSASAEAAKAHLAGKNMQPSQLSYSTSPSPSLPINPGTSHKHH	240

Db 181 YIVPHGGHYIIPKSDLSASELAARAHLAGKNQPSQSLSTPSPILNPGTSHEKH 240
Qy 241 EDCYGFDAIRIIAEDSGFVMSHGDHNYFFKKDLTEQIKAAQKHLSEVKTSHNGLDS 300
Db 241 EDCYGFDAIRIIAEDSGFVMSHGDHNYFFKKDLTEQIKAAQKHLSEVKTSHNGLDS 300
Qy 301 LSSHEQDYPNAKEMKDLKKIEEKIAGIMQYGVRESIVVNVKKNALIIYPHGDHHDAD 360
Db 301 LSSHEQDYPNAKEMKDLKKIEEKIAGIMQYGVRESIVVNVKKNALIIYPHGDHHDAD 360
Qy 361 PIDCHKPVGCHSHSNVELFKPEGVAKEGKNVYTGEEELTNVNLKSTFNQNFLLA 420
Db 361 PIDCHKPVGCHSHSNVELFKPEGVAKEGKNVYTGEEELTNVNLKSTFNQNFLLA 420
Qy 421 NGKRVSFSPPELEKKGILNMLVKLITPDGKVLKVSFGVGVGNIAFELDQYLP 480
Db 421 NGKRVSFSPPELEKKGILNMLVKLITPDGKVLKVSFGVGVGNIAFELDQYLP 480
Qy 481 GQPKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPHAGDTYLRVNPQFAVPKGT 540
Db 481 GQPKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPHAGDTYLRVNPQFAVPKGT 540
Qy 541 DALVRVDFEHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTMANAYLDNQSTY 600
Db 541 DALVRVDFEHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTMANAYLDNQSTY 600
Qy 601 IVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLGSETGNSTN 660
Db 601 IVEVPILEKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLGSETGNSTN 660
Qy 661 STLBEVPTDVPQEKVAKFAESYGMKLENVLFNNMDGTIELYLPSSGEVKKNNLLIS 720
Db 661 STLBEVPTDVPQEKVAKFAESYGMKLENVLFNNMDGTIELYLPSSGEVKKNNLLIS 720
Qy 721 QGNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNENSTDGMLNPEGNVGS 780
Db 721 QGNGENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNENSTDGMLNPEGNVGS 780
Qy 781 DPMLDLSALEAPADVPQEKLEKFTASYGGLDSVIFNMDGTIELRPSGEVKKNNLLIS 840
Db 781 DPMLDLSALEAPADVPQEKLEKFTASYGGLDSVIFNMDGTIELRPSGEVKKNNLLIS 840

RESULT 2

US-09-884-465A-6
; Sequence 6, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 035190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1039
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-6

Query Match 97.38; Score 4276.5; DB 9; Length 1039;
Best Local Similarity 82.38; Pred. No. 1.4e-245;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Caps 1;

Qy 1 CAYALNQRHROENKDNRRVSYVDGSSQSKSENLTPOVSOKEGIAEQIVIKITDQGYV 60
Db 21 CAYALNQRHROENKDNRRVSYVDGSSQSKSENLTPOVSOKEGIAEQIVIKITDQGYV 80
Qy 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADADIVNEVKGGYIIKVDGKYVYVLKD 120
Db 81 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADADIVNEVKGGYIIKVDGKYVYVLKD 140
Qy 121 RAHADNVTKDEINRQKQEHVKDNKNSVAVARSQGYTTNDGYVFNPAIDTIEDTGN 180
Db 141 RAHADNVTKDEINRQKQEHVKDNKNSVAVARSQGYTTNDGYVFNPAIDTIEDTGN 200
Qy 181 YIVPHGGHYHYIPKSDLSASELAARAHLAGKNQPSQSLSTPSPILNPGTSHEKH 223
Db 201 YIVPHGGHYHYIPKSDLSASELAARAHLAGKNQPSQSLSTPSPILNPGTSHEKH 260
Qy 224 ----- 223
Db 261 PANKSENLSQLLKELYDSPAQRYSSEDLVDFPAKIIISRTPNGVAIPHGDHYHFIYSK 320
Qy 224 ----- 223
Db 321 LSALBEKTAHVPIISGTGSTVSTNAKPNEVSSLSLSSNPSSLTTSKELSSASDGYFN 380
Qy 224 -----TPSPSLPINPOTSHEKHEED 243
Db 381 PKDIVEETATAVIVRHGDHFHYIPKSNQIQGPTLPNNSLATPSPSLPINPOTSHEKHEED 440
Qy 244 GYGFANRIIAEDSGFVMSHGDHNYFFKKDLTEQIKAAQKHLSEVKTSHNGLDSLS 303
Db 441 GYGFANRIIAEDSGFVMSHGDHNYFFKKDLTEQIKAAQKHLSEVKTSHNGLDSLS 500
Qy 304 HQDYPXNAKEMKDLKKIEEKIAGIMQYGVRESIVVNVKKNALIIYPHGDHHDADID 363
Db 501 HQDYPXNAKEMKDLKKIEEKIAGIMQYGVRESIVVNVKKNALIIYPHGDHHDADID 560
Qy 364 EHKPVGIGHSHSNYELFKPEGVAKEGKNVYTGEEELTNVNLKSTFNQNFLLA 423
Db 561 EHKPVGIGHSHSNYELFKPEGVAKEGKNVYTGEEELTNVNLKSTFNQNFLLA 620
Qy 424 KRVSFSPPELEKKGILNMLVKLITPDGKVLKVSFGVGVGNIAFELDQYLPQGT 483
Db 621 KRVSFSPPELEKKGILNMLVKLITPDGKVLKVSFGVGVGNIAFELDQYLPQGT 680
Qy 484 FKYTIAKDYPEVSYDGTFTVPTSLAYKMASQTIYFPHAGDTYLRVNPQFAVPKGTAL 543
Db 681 FKYTIAKDYPEVSYDGTFTVPTSLAYKMASQTIYFPHAGDTYLRVNPQFAVPKGTAL 740
Qy 544 VRVDFEHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTMANAYLDNQSTYIVE 603
Db 741 VRVDFEHGNAYLENNYKVGEIKLPIPKLNOGTTTAGNKIPVTMANAYLDNQSTYIVE 800
Qy 604 VPILKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLGSETGNSTN 663
Db 801 VPILKENQTDKPSILPOFKRNKAQENSKLDEKVEEPTSEKVEKEKLGSETGNSTN 860
Qy 664 EEPVTVDPQEKVAKFAESYGMKLENVLFNNMDGTIELYLPSSGEVKKNNMDTGEAPOGN 723
Db 861 EEPVTVDPQEKVAKFAESYGMKLENVLFNNMDGTIELYLPSSGEVKKNNMDTGEAPOGN 920
Qy 724 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNENSTDGMLNPEGNVGSDFM 783
Db 921 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNENSTDGMLNPEGNVGSDFM 980
Qy 784 LLSALEAPADVPQEKLEKFTASYGGLDSVIFNMDGTIELRPSGEVKKNNL 837
Db 981 LLSALEAPADVPQEKLEKFTASYGGLDSVIFNMDGTIELRPSGEVKKNNL 1034

RESULT 3

US-09-884-465A-379
; Sequence 379, Application US/09884465A
; Publication No. US20030077293A1

```

: GENERAL INFORMATION:
: APPLICANT: Shire Biochem, Inc.
: APPLICANT: Hamel, Josee
: APPLICANT: Brodeur, Bernard
: APPLICANT: Martin, Denis
: APPLICANT: Charland, Nathalie
: APPLICANT: Ouellet, Catherine
: TITLE OF INVENTION: Streptococcus Antigens
: FILE REFERENCE: 055190-0044
: CURRENT APPLICATION NUMBER: US/09/884,465A
: CURRENT FILING DATE: 2001-06-20
: PRIOR APPLICATION NUMBER: 60/212,683
: PRIOR FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 384
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 379
: LENGTH: 1152
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Unknown Organism
: NAME/KEY: MISC_FEATURE
: LOCATION: (1)..(1)
: OTHER INFORMATION: xaa = Methionine or nothing
: NAME/KEY: MISC_FEATURE
: LOCATION: (344)..(344)
: OTHER INFORMATION: xaa = Glycine or nothing
: NAME/KEY: MISC_FEATURE
: LOCATION: (345)..(345)
: OTHER INFORMATION: xaa = Proline or nothing
: US-09-884-465A-379

Query Match          77.4%; Score 3402.5; DB 9; Length 1152;
Best Local Similarity 79.6%; Pred. No. 1e-193;
Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11;

QY 10 SQENKNNRVSYVDGQS--SOKSENL-----TPDOVSQKEGTOAEQI-VIKIT 55
DB 355 SSTASDNNQTSVAKGSTSPANKNSLQKELXDPSAQRYSDGLVFPDAKIISRT 414
QY 56 DQGYVTSBGHDHYHYNGKYPYDALFS-EELLMKDPNYQLKDADIVNEVKGVIKVDGKY 114
DB 415 PNGVAIPHGDHYHF-----IPYKLSALEEKTAR-----MVPISGT- 450
QY 115 YVYLKDAADNADNRVTKDEINROKQEHVKNEKNSVAVARSQGRYTT-----NDGYV 167
DB 451 -----GSTVST-----NAKDNVYSSLSGLSSNPSSLTTSKELSSASDGYI 491
QY 168 FNPADIIDTGNAYIVPHGCHYHYTPKSDLSASELAAGKAGNMOPSQLSYS-STPS 226
DB 492 FNPKDIVEETATAYIVRHGDHFPYTPKSNQIG-----QTLFNNSLATPS 536
QY 227 PSLPINPGTSHKEHEDGCGFDANRTIAEDSGFVNSHGDHNNHYFFKDLTEEQIKAAQK 286
DB 537 PSLPINPGTSHKEHEDGCGFDANRTIAEDSGFVNSHGDHNNHYFFKDLTEEQIKAAQK 596
QY 287 HLEEVKTSUNGDLSSSHQDDYPSNAKEMKDKLTKTEEKIAGIMKOYGVKRESIVVYNEK 346
DB 597 HLEEVKTSUNGDLSSSHQDDYPSNAKEMKDKLTKTEEKIAGIMKOYGVKRESIVVYNEK 656
QY 347 NAIYIPHGDDHHPIDIDEKPKVIGHSNHYELFKPEEGVAKKEGKVKYVTGEELTNVYNL 406
DB 657 NAIYIPHGDDHHPIDIDEKPKVIGHSNHYELFKPEEGVAKKEGKVKYVTGEELTNVYNL 716
QY 407 LKNSIFNNQNFILANGQKVSFSFPELEKLGILNMLVKLIIPDQKVLKVSCKVFGGV 466
DB 717 LKNSIFNNQNFILANGQKVSFSFPELEKLGILNMLVKLIIPDQKVLKVSCKVFGGV 776
QY 467 GNIAFELDQPLPGOTFKYTIASKDYPVSYDGTFTVPTSLAYKMASOTIYPFHAGDT 526
DB 777 GNIAFELDQPLPGOTFKYTIASKDYPVSYDGTFTVPTSLAYKMASOTIYPFHAGDT 836
QY 527 YLRVNPQFAVPKGTDALVRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPV 586

: GENERAL INFORMATION:
: APPLICANT: Shire Biochem, Inc.
: APPLICANT: Hamel, Josee
: APPLICANT: Brodeur, Bernard
: APPLICANT: Martin, Denis
: APPLICANT: Charland, Nathalie
: APPLICANT: Ouellet, Catherine
: TITLE OF INVENTION: Streptococcus Antigens
: FILE REFERENCE: 055190-0044
: CURRENT APPLICATION NUMBER: US/09/884,465A
: CURRENT FILING DATE: 2001-06-20
: PRIOR APPLICATION NUMBER: 60/212,683
: PRIOR FILING DATE: 2000-06-20
: NUMBER OF SEQ ID NOS: 384
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 381
: LENGTH: 1238
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Unknown Organism
: NAME/KEY: MISC_FEATURE
: LOCATION: (1)..(1)
: OTHER INFORMATION: xaa = Methionine or nothing
: NAME/KEY: MISC_FEATURE
: LOCATION: (430)..(430)
: OTHER INFORMATION: xaa = Glycine or nothing
: NAME/KEY: MISC_FEATURE
: LOCATION: (431)..(431)
: OTHER INFORMATION: xaa = Proline or nothing
: US-09-884-465A-381

Query Match          77.4%; Score 3402.5; DB 9; Length 1238;
Best Local Similarity 79.6%; Pred. No. 1.1e-193;
Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11;

QY 10 SQENKNNRVSYVDGQS--SOKSENL-----TPDOVSQKEGTOAEQI-VIKIT 55
DB 441 SSTASDNNQTSVAKGSTSPANKNSLQKELXDPSAQRYSDGLVFPDAKIISRT 500
QY 56 DQGYVTSBGHDHYHYNGKYPYDALFS-EELLMKDPNYQLKDADIVNEVKGVIKVDGKY 114
DB 501 PNGVAIPHGDHYHF-----IPYKLSALEEKTAR-----MVPISGT- 536
QY 115 YVYLKDAADNADNRVTKDEINROKQEHVKNEKNSVAVARSQGRYTT-----NDGYV 167
```

Db 537 -----GTVST-----NAPNEVSSLSLSSPSSLTTSKELSSASDGYI 577
QY 168 FNPADIIEDTGNAYIVPHGCHYHYIPKSDLSASELAHAKAHLAGKMMQPSOLSYS-STPS 226
Db 578 FNPKDIVEETATAYIVRHGDHFIYIPKSNQIG-----OPTLPNNSLATPS 622
QY 227 PSLPINTGTSHEKHEEDGCGFDANRIIADDESGFVMSHGDHNYFFKDLTDEEQIKAAQK 286
Db 623 PSLPINTGTSHEKHEEDGCGFDANRIIADDESGFVMSHGDHNYFFKDLTDEEQIKAAQK 682
QY 287 HLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKIEKIAIMQYGVKRESIVVNEK 346
Db 683 HLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKIEKIAIMQYGVKRESIVVNEK 742
QY 347 NAIYYPHGDHHDADPIDEHKPVGIGHSHSNYELFPEEGVAKKGNKYVTGELNMYNL 406
Db 743 NAIYYPHGDHHDADPIDEHKPVGIGHSHSNYELFPEEGVAKKGNKYVTGELNMYNL 802
QY 407 LKNSTFNQNTLANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKESGKVFGEV 466
Db 803 LKNSTFNQNTLANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKESGKVFGEV 862
QY 467 GNIANFELDQPYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDT 526
Db 863 GNIANFELDQPYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDT 922
QY 527 YLRVNPQFAVPKGTDALVRVDFEHGNAYLNNYKVEIKLPIPKLNQGTTRTAGNKIPV 586
Db 923 YLRVNPQFAVPKGTDALVRVDFEHGNAYLNNYKVEIKLPIPKLNQGTTRTAGNKIPV 982
QY 587 TFMANAYLDNQSTYIIEVPILEKENQTDKPSILPOLKRNKAQENSKLDEKVEEPTSEKV 646
Db 983 TFMANAYLDNQSTYIIEVPILEKENQTDKPSILPOLKRNKAQENSKLDEKVEEPTSEKV 1042
QY 647 EKEKLSGTSTNSNLTLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSE 706
Db 1043 EKEKLSGTSTNSNLTLEEVPTVDPQVKVAKFAESYGMKLENVLFNMDGTIELYLPSE 1102
QY 707 VIKKNADFTGEAPQNGENKPSNGKVTGTVENOPTENKPADSLPEAPNEKPVKPSNS 766
Db 1103 VIKKNADFTGEAPQNGENKPSNGKVTGTVENOPTENKPADSLPEAPNEKPVKPSNS 1162
QY 767 TDNGMLNPEGNVGSDPMDLSALSEAPAVDPVQEKLEKFTASYGLGLDSVTFNMDGTIELR 826
Db 1163 TDNGMLNPEGNVGSDPMDLPALSEAPAVDPVQEKLEKFTASYGLGLDSVTFNMDGTIELR 1222
QY 827 LPSGEVIKKNL 837
Db 1223 LPSGEVIKKNL 1233

RESULT 5
US-09-884-465A-382
; Sequence 382, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,583
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 382
; LENGTH: 1365

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (557)..(557)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: Xaa = Proline or nothing
; US-09-884-465A-382

Query Match 77.4%; Score 3402.5; DB 9; Length 1365;
Best Local Similarity 79.6%; Pred. No. 1.2e-193;
Matches 677; Conservative 31; Mismatches 62; Indels 81; Gaps 11;

QY 10 SQENKDNRRVSVVDGSSQ--SOKSENL-----TPDQVSQKEGIOAEQT-VIKIT 55
Db 568 SSTASDNTQSVAKGSTSKPANKSENQSLKELYDSPSAQRYSESGLYDFDPAKIIIRT 627
QY 56 DQGYVTSYSHGDHNYNGKVPYDALFS-EELLMKDPNYOLKADADIVNEVGKGIYIKVDGKY 114
Db 628 PNGVAIPGHGDHYHF----IPYSKLSALEEKIAR-----MVPISGT- 663
QY 115 VYVLKDAHADNVRTKDEINPQKOEHVKNDEKVNNAVARSQGYTT-----NDGYV 167
Db 664 -----GSTVST-----NAKPNEVSVSLGSLSSNPSLTSKELSSASDGYI 704
QY 168 FNPADIIEDTGNAYIVPHGCHYHYIPKSDLSASELAHAKAHLAGKMMQPSOLSYS-STPS 226
Db 705 FNPKDIVEETATAYIVRHGDHFIYIPKSNQIG-----OPTLPNNSLATPS 749
QY 227 PSLPINTGTSHEKHEEDGCGFDANRIIADDESGFVMSHGDHNYFFKDLTDEEQIKAAQK 286
Db 750 PSLPINTGTSHEKHEEDGCGFDANRIIADDESGFVMSHGDHNYFFKDLTDEEQIKAAQK 809
QY 287 HLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKIEKIAIMQYGVKRESIVVNEK 346
Db 810 HLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDLKDKIEKIAIMQYGVKRESIVVNEK 869
QY 347 NAIYYPHGDHHDADPIDEHKPVGIGHSHSNYELFPEEGVAKKGNKYVTGELNMYNL 406
Db 870 NAIYYPHGDHHDADPIDEHKPVGIGHSHSNYELFPEEGVAKKGNKYVTGELNMYNL 929
QY 407 LKNSTFNQNTLANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKESGKVFGEV 466
Db 930 LKNSTFNQNTLANGOKRVSFSPPELEKKGINMLVKLITPDGKVLKESGKVFGEV 989
QY 467 GNIANFELDQPYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDT 526
Db 990 GNIANFELDQPYLPQGTFTYIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDT 1049
QY 527 YLRVNPQFAVPKGTDALVRVDFEHGNAYLNNYKVEIKLPIPKLNQGTTRTAGNKIPV 586
Db 1050 YLRVNPQFAVPKGTDALVRVDFEHGNAYLNNYKVEIKLPIPKLNQGTTRTAGNKIPV 1109
QY 587 TFMANAYLDNQSTYIIEVPILEKENQTDKPSILPOLKRNKAQENSKLDEKVEEPTSEKV 646
Db 1110 TFMANAYLDNQSTYIIEVPILEKENQTDKPSILPOLKRNKAQENSKLDEKVEEPTSEKV 1169
QY 647 EKEKLSGTSTNSNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSE 706
Db 1170 EKEKLSGTSTNSNLTLEEVPTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSE 1229
QY 707 VIKKNADFTGEAPQNGENKPSNGKVTGTVENOPTENKPADSLPEAPNEKPVKPSNS 766
Db 1230 VIKKNADFTGEAPQNGENKPSNGKVTGTVENOPTENKPADSLPEAPNEKPVKPSNS 1289
QY 767 TDNGMLNPEGNVGSDPMDLSALSEAPAVDPVQEKLEKFTASYGLGLDSVTFNMDGTIELR 826
Db 1223 LPSGEVIKKNL 1233


```

RESULT 9
US-09-884-465A-372
; Sequence 372, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884, 465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 372
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-372

Query Match 73.2%; Score 3218; DB 9; Length 906;
Best Local Similarity 73.3%; Pred. No. 6.3e-183;
Matches 656; Conservative 35; Mismatches 84; Indels 120; Gaps 12;

QY 25 SOSQSSENLTPDQVSOKEGIAE-----QIVKITDQG 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 45 AQAYAKEKGLTPSTDHQDSGNTAKGAEALYNRVKAAKVPDLDRMPYNLQYTVVEKNGS 104
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 59 YVTSHGDIHYHYNGKVPYDALFSEILLMKDPNQLKADIVNEVKGYIIVKDGKYYVYL 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 105 LIIPHYDIHY- -NIKFW- --FDGLYEAAPKGSLED-----LLATKYVVEP 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 KDAAHADNVRTKDEINRQKHVDKNEKNSNVAVARSGQRYTTNDGVYENPADIIDTG 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 RNA-SDHVRK- ---NKADQDSKPEDKDEHV- ---SETPHPSD-----EKEN 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 179 NAVIVPGGHYHYIPKSDLSASELAAA- ---KAHLG----- 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 HAGLNPADNL-YKPSPTDTEETEEAEADTIDEAIPGTPSIQNAMETLTGLKSSLLGT 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 212 -----KKNQPSQL-----SYSTSPSPSPINPGTSHEKHEE 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 KDNNTISAEVDSLALLKESQAPIQGQPGIOTPLPNNSLATSPSPINPGTSHEKHEE 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 DGYGFADNRLLIAEDSGFVMSHDHNNHYFFKDLTTEQIKAAQKHLEEVKTSNGLDLS 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 DGYGFADNRLLIAEDSGFVMSGSDSNHYFFKDLTTEQIKAAQKHLEEVKTSNGLDLS 366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 303 SHEQDPSNAKEMKDDKKTEEKIAGITMKQYGVKRESIVVNEKNAIIYPHGDHHDADI 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 367 SHEQDPSNAKEMKDDKKTEEKIAGITMKQYGVKRESIVVNEKNAIIYPHGDHHDADI 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 363 DEHKPVGIGHSHSNYELFKPEEGVAKKGNKVTGSELITNVNLLKNSTNNQNFTLANG 422
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 427 DEHKPVGIGHSHSNYELFKPEEGVAKKGNKVTGSELITNVNLLKNSTNNQNFTLANG 486
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 423 QKRVSRFPPELKKKGINMLVLIITPDGKVLKVSQKVFGEVGNITANFELDQYLPQG 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 487 QKRVSRFPPELKKKGINMLVLIITPDGKVLKVSQKVFGEVGNITANFELDQYLPQG 546
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 483 TKYITASKDYPEVSYDGTFTVPTSLAYKASOTIFYPPHAGDTYLRVNPQFAVPKGTDA 542
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 547 TKYITASKDYPEVSYDGTFTVPTSLAYKASOTIFYPPHAGDTYLRVNPQFAVPKGTDA 606
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 543 LVRVDFEFGHNAVLENNYKVGETIKLPILQNGQTRTAGNKIPVTFMANAYLDNQSYIV 602
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 607 LVRVDFEHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTMANAYLDNQSTYIV 666
Qy 603 EVPILEXENQTDKPSILPQPKRKAQENSKLDEKVEBPKTSEKVEKEKLTSETGNSNST 662
Db 667 EVPILEXENQTDKPSILPQPKRKAQENSKLDEKVEBPKTSEKVEKEKLTSETGNSNST 726
Qy 663 LEEVPTVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVTKKNMADFTGEAPQG 722
Db 727 LEEVPTVDPVOEKVAKFAESYGMKLENVLFNMDGTIELYLPSPGEVTKKNMADFTGEAPQG 786
Qy 723 NGENKPSGKSVSTGTVENOPTENKPADSLPEAPNEKPVKPENSTONGMLNPGCVGSDP 782
Db 787 NGENKPSGKSVSTGTVENOPTENKPADSLPEAPNEKPVKPENSTONGMLNPGCVGSDP 846
Qy 783 MLDSALEEAPAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVTKKNL 837
Db 847 MLDPALPEAPVAVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVTKKNL 901

RESULT 10
US-09-884-465A-334
; Sequence 334, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 334
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-334

```

[illegible]

APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 361
LENGTH: 633
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-361

Query Match 72.6%; Score 3193.5; DB 9; Length 633;
Best Local Similarity 98.4%; Pred. No. 1.1e-181;
Matches 614; Conservative 3; Mismatches 6; Indels 1; Gaps 1;
QY 215 QPSQSYS-STPSPSLPINPGTSHKHEEDGYGFDANRIIADESGFVMSHGDHNYFFK 273
DB 5 QPTLPNNSLATPSPSLPINPGTSHKHEEDGYGFDANRIIADESGFVMSHGDHNYFFK 64
QY 274 KDLTEEQIKAAQKHLLEVKTSHNGLDLSLSSHEODYPSNAKEMKDLKKIEEKIAGIMKQY 333
DB 65 KDLTEEQIKAAQKHLLEVKTSHNGLDLSLSSHEODYPSNAKEMKDLKKIEEKIAGIMKQY 124
QY 334 GVKRESIVVKNENKNAIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPEEGVAKKEGK 393
DB 125 GVKRESIVVKNENKNAIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPEEGVAKKEGK 184
QY 394 VYTGEELTNVNLKNSTFNQNTLANGOKRVSFSPFPELEKLGINMLVKLITPDGKV 453
DB 185 VYTGEELTNVNLKNSTFNQNTLANGOKRVSFSPFPELEKLGINMLVKLITPDGKV 244
QY 454 LEKVGKVFGEVGNIANFELDQPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMA 513
DB 245 LEKVGKVFGEVGNIANFELDQPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMA 304
QY 514 SOTIYFPHAGDTYLRVNPQFAPVPGTDLALRVDFDEFGHGNAYLENNYKVEIKLPIPKLN 573
DB 305 SOTIYFPHAGDTYLRVNPQFAPVPGTDLALRVDFDEFGHGNAYLENNYKVEIKLPIPKLN 364
QY 574 QGTTTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKENQTDKPSILPQFKNKKAQENSKL 633
DB 365 QGTTTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKENQTDKPSILPQFKNKKAQENSKL 424
QY 634 DEKVEEPTSEKVEKEKLSGTNSTSNSTLEEVPTDVPQEKVAKFAESYGMKLENNVLEN 693
DB 425 DEKVEEPTSEKVEKEKLSGTNSTSNSTLEEVPTDVPQEKVAKFAESYGMKLENNVLEN 484
QY 694 MDGTIELYLPAGEVKKMAADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLP 753
DB 485 MDGTIELYLPAGEVKKMAADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLP 544
QY 754 EAPNEKVPKPNSTNGMLNPEGNVGSDPMLDLSALEAPAVDPVQEKLEKFTASYGLGLD 813
DB 545 EAPNEKVPKPNSTNGMLNPEGNVGSDPMLDLSALEAPAVDPVQEKLEKFTASYGLGLD 604
QY 814 SVIFNMDGTIELRLPSGEVIKKNL 837
DB 605 SVIFNMDGTIELRLPSGEVIKKNL 628

RESULT 15
US-09-884-465A-362
Sequence 362, Application US/09884465A
Publication No US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee

APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charliet, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 362
LENGTH: 633
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-362

Query Match 72.6%; Score 3193.5; DB 9; Length 633;
Best Local Similarity 98.4%; Pred. No. 1.1e-181;
Matches 614; Conservative 3; Mismatches 6; Indels 1; Gaps 1;
QY 215 QPSQSYS-STPSPSLPINPGTSHKHEEDGYGFDANRIIADESGFVMSHGDHNYFFK 273
DB 5 QPTLPNNSLATPSPSLPINPGTSHKHEEDGYGFDANRIIADESGFVMSHGDHNYFFK 64
QY 274 KDLTEEQIKAAQKHLLEVKTSHNGLDLSLSSHEODYPSNAKEMKDLKKIEEKIAGIMKQY 333
DB 65 KDLTEEQIKAAQKHLLEVKTSHNGLDLSLSSHEODYPSNAKEMKDLKKIEEKIAGIMKQY 124
QY 334 GVKRESIVVKNENKNAIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPEEGVAKKEGK 393
DB 125 GVKRESIVVKNENKNAIYPHGDHHDADIDEHKPVGIGHSHSNYELFKPEEGVAKKEGK 184
QY 394 VYTGEELTNVNLKNSTFNQNTLANGOKRVSFSPFPELEKLGINMLVKLITPDGKV 453
DB 185 VYTGEELTNVNLKNSTFNQNTLANGOKRVSFSPFPELEKLGINMLVKLITPDGKV 244
QY 454 LEKVGKVFGEVGNIANFELDQPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMA 513
DB 245 LEKVGKVFGEVGNIANFELDQPYLPQGTFKYTIASKDYPEVSYDGTFTVPTSLAYKMA 304
QY 514 SOTIYFPHAGDTYLRVNPQFAPVPGTDLALRVDFDEFGHGNAYLENNYKVEIKLPIPKLN 573
DB 305 SOTIYFPHAGDTYLRVNPQFAPVPGTDLALRVDFDEFGHGNAYLENNYKVEIKLPIPKLN 364
QY 574 QGTTTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKENQTDKPSILPQFKNKKAQENSKL 633
DB 365 QGTTTAGNKIPVTFMANAYLDNQSTYIIVEPILKEKENQTDKPSILPQFKNKKAQENSKL 424
QY 634 DEKVEEPTSEKVEKEKLSGTNSTSNSTLEEVPTDVPQEKVAKFAESYGMKLENNVLEN 693
DB 425 DEKVEEPTSEKVEKEKLSGTNSTSNSTLEEVPTDVPQEKVAKFAESYGMKLENNVLEN 484
QY 694 MDGTIELYLPAGEVKKMAADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLP 753
DB 485 MDGTIELYLPAGEVKKMAADFTGEAPQNGENKPSNGKSVSTGTVENQPTENKPADSLP 544
QY 754 EAPNEKVPKPNSTNGMLNPEGNVGSDPMLDLSALEAPAVDPVQEKLEKFTASYGLGLD 813
DB 545 EAPNEKVPKPNSTNGMLNPEGNVGSDPMLDLSALEAPAVDPVQEKLEKFTASYGLGLD 604
QY 814 SVIFNMDGTIELRLPSGEVIKKNL 837
DB 605 SVIFNMDGTIELRLPSGEVIKKNL 628

Search completed: May 13, 2003, 13:59:01
Job time : 28.7625 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:49:42 : Search time 43.5434 Seconds
(without alignments)
3118.326 Million cell updates/sec

Title: US-09-471-255-55
Perfect score: 5315
Sequence: 1 CAYALNQHRSEKNNRVS.....IELRLPSGEVTKKLSDFIA 1019

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

```

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	5315	100.0	1019 21	AA12722 Streptococcus pneu
2	5315	100.0	1019 21	AA12751 Streptococcus pneu
3	5315	100.0	1019 23	AAU84021 Truncated variant
4	5315	100.0	1039 21	AA12715 Streptococcus pneu
5	5315	100.0	1039 23	AAU75932 Streptococcus pneu
6	5297	99.7	1019 21	AA12750 Streptococcus pneu
7	5291	99.5	1019 21	AA12748 Streptococcus pneu
8	5291	99.5	1019 21	AA12749 Streptococcus pneu
9	5291	99.5	1019 21	AA12752 Streptococcus pneu
10	5291	99.5	1019 21	AA12753 Streptococcus pneu

11	4325	81.4	1238	23	AAU84056	S. pneumoniae deri
12	4276.5	80.5	840	21	AA12721	Streptococcus pneu
13	4276.5	80.5	840	23	AAU76151	Streptococcus pneu
14	4263.5	80.2	1378	23	AAU84053	S. pneumoniae deri
15	4261.5	80.2	1365	23	AAU84057	S. pneumoniae deri
16	4244	79.8	1152	23	AAU84054	S. pneumoniae deri
17	4196	78.9	807	23	AAU83997	Truncated variant
18	4090	77.0	780	21	AA12744	Streptococcus pneu
19	3559	67.0	679	23	AAU84092	Truncated variant
20	3384	63.7	907	23	AAU84061	S. pneumoniae deri
21	3381	63.6	644	23	AAU83996	Truncated variant
22	3375	63.5	907	23	AAU84065	S. pneumoniae deri
23	3366	63.3	907	23	AAU84063	S. pneumoniae deri
24	3362	63.3	901	23	AAU84087	S. pneumoniae deri
25	3345.5	62.9	902	23	AAU84076	S. pneumoniae deri
26	3338	62.8	901	23	AAU84068	S. pneumoniae deri
27	3329	62.6	901	23	AAU84072	S. pneumoniae deri
28	3329	62.6	901	23	AAU84074	S. pneumoniae deri
29	3327.5	62.6	902	23	AAU84078	S. pneumoniae deri
30	3325	62.6	895	23	AAU84070	S. pneumoniae deri
31	3323.5	62.5	896	23	AAU84080	S. pneumoniae deri
32	3306	62.2	632	23	AAU83998	Truncated variant
33	3297	62.0	632	23	AAU84002	Truncated variant
34	3297	62.0	632	23	AAU84006	Truncated variant
35	3297	62.0	632	23	AAU84013	Truncated variant
36	3297	62.0	632	23	AAU84014	Truncated variant
37	3288.5	61.9	896	23	AAU84077	S. pneumoniae deri
38	3288	61.9	632	23	AAU84003	Truncated variant
39	3288	61.9	632	23	AAU84004	Truncated variant
40	3288	61.9	632	23	AAU84009	Truncated variant
41	3288	61.9	632	23	AAU84012	Truncated variant
42	3288	61.9	901	23	AAU84066	S. pneumoniae deri
43	3288	61.9	907	23	AAU84060	S. pneumoniae deri
44	3288	61.9	907	23	AAU84064	S. pneumoniae deri
45	3287	61.8	632	23	AAU84015	Truncated variant

ALIGNMENTS

RESULT 1
AA12722
ID AA12722 standard; Protein; 1019 AA.
XX AA12722;
XX AC
XX AA12722;
DT 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae BVH-3M protein antigen SEQ ID NO:55.
XX KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX OS Streptococcus pneumoniae.
XX FN WO200039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA01218.
XX PR 23-DEC-1998; 98US-0113800.
XX FA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hanel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX DR WPI: 2000-452397/39.
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteraemia and/or pneumonia

PS	Claim 18; Fig 20; 106pp; English.	
XX	The present invention describes nucleic acids (I) encoding protein	
CC	antigens (II) from Streptococcus pneumoniae. The protein antigens	
CC	have bactericidal activity. The nucleic acids, encoding the protein	
CC	antigens, may be used for the recombinant production of the proteins	
CC	they encode. The protein antigens may then be used as vaccines for the	
CC	prevention and treatment of Streptococcal infections in mammals	
CC	(especially humans) which result in, e.g. meningitis, otitis media,	
CC	bacteraemia and/or pneumonia. The present sequence represents the	
CC	S. pneumoniae BVH-3M protein antigen.	
XX		
SQ	Sequence 1019 AA:	
	Query Match 100.0%; Score 5315; DB 21; Length 1019;	
	Best Local Similarity 100.0%; Pred. No. 1.3e-314;	
	Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 CAYALNQRSQENKNNRYSYVDGSSQSKSENLPDQVSQKREGIQAEQIVIKITDQGV 60	
DB	1 CAYALNQRSQENKNNRYSYVDGSSQSKSENLPDQVSQKREGIQAEQIVIKITDQGV 60	
QY	61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQKADIVNEVKGGYIIRKVDGKYVYLKD 120	
DB	61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQKADIVNEVKGGYIIRKVDGKYVYLKD 120	
QY	121 AAHADNVKDEINROKQEHVXDNEKYNVAVARSQGYTTNDGYVFPADIIEDTGA 180	
DB	121 AAHADNVKDEINROKQEHVXDNEKYNVAVARSQGYTTNDGYVFPADIIEDTGA 180	
QY	181 YIVPHGHHYIPKSDLSASELAARAHLAGNMOPSQLSYSTASDNNNTQSVAGSTSK 240	
DB	181 YIVPHGHHYIPKSDLSASELAARAHLAGNMOPSQLSYSTASDNNNTQSVAGSTSK 240	
QY	241 PANKSNLOSLLKELYDSPAQRYSSEGLVDPAPKIIISRTNGVAIPHGDIHPIPIYSK 300	
DB	241 PANKSNLOSLLKELYDSPAQRYSSEGLVDPAPKIIISRTNGVAIPHGDIHPIPIYSK 300	
QY	301 LSALEKIRARVPISGTGSTVSTNAKPNEVSWSLGSLSNPSSLTSSKLSASDGYIFN 360	
DB	301 LSALEKIRARVPISGTGSTVSTNAKPNEVSWSLGSLSNPSSLTSSKLSASDGYIFN 360	
QY	361 PKDIVEETATAYIVRHGDHFHYIPKSNQIQGPTLPNNSLATPSPLINPGTSHEKHEED 420	
DB	361 PKDIVEETATAYIVRHGDHFHYIPKSNQIQGPTLPNNSLATPSPLINPGTSHEKHEED 420	
QY	421 GYGFDAIRIIAEDSGFVMSHGDNHYFFKDLTEBQIRAAOKHLEEVKTSNGLDLSLS 480	
DB	421 GYGFDAIRIIAEDSGFVMSHGDNHYFFKDLTEBQIRAAOKHLEEVKTSNGLDLSLS 480	
QY	481 HEDYFGNAKEMKDLKKIBEKIAGINKOYGVKRESIVVYVNEKNAIYPHGDHHDADPID 540	
DB	481 HEDYFGNAKEMKDLKKIBEKIAGINKOYGVKRESIVVYVNEKNAIYPHGDHHDADPID 540	
QY	541 EHRPVGIGSHSHNYELFKPEGVAKGKNVYTGEBELTNVWLLKNTFNNOFTLANGQ 600	
DB	541 EHRPVGIGSHSHNYELFKPEGVAKGKNVYTGEBELTNVWLLKNTFNNOFTLANGQ 600	
QY	601 KRVSFFPPPELEKLGINMLVLIITPDGKLVKYSKVFEGVGNIANFELDOPYLPQGT 660	
DB	601 KRVSFFPPPELEKLGINMLVLIITPDGKLVKYSKVFEGVGNIANFELDOPYLPQGT 660	
QY	661 FKYTIAKQVPEVSYDGTFTVPTSLAYKMASQITFPFHAGDTYLRVNPQFAPVKGTDAL 720	
DB	661 FKYTIAKQVPEVSYDGTFTVPTSLAYKMASQITFPFHAGDTYLRVNPQFAPVKGTDAL 720	
QY	721 VRVDFEFGHAYLNNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 780	
DB	721 VRVDFEFGHAYLNNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSYIVE 780	
QY	781 VPILEKENQDKPSILPQFKRNKAQENSKLDEKVEBPKTSEKVEKEKLSGTGNSNSTL 840	
DB	781 VPILEKENQDKPSILPQFKRNKAQENSKLDEKVEBPKTSEKVEKEKLSGTGNSNSTL 840	


```
QY 121 AAHADNVRTKDEINROKQEHVKDKNEKNSVAVARSGQRYTTNDGYVFNPAIDIEDTGN 180
Db 121 AAHADNVRTKDEINROKQEHVKDKNEKNSVAVARSGQRYTTNDGYVFNPAIDIEDTGN 180
QY 181 YIVPHGCHYHYIPKSDLSASELAHAKAHLAGKNQPSQLSYSTASDNTQSVAKGSTSK 240
Db 181 YIVPHGCHYHYIPKSDLSASELAHAKAHLAGKNQPSQLSYSTASDNTQSVAKGSTSK 240
QY 241 PANKSENLOSLKELYDPSAQRYSESDGLVFPDAKIISRTPNGVAIPHGDHYHFIPYSK 300
Db 241 PANKSENLOSLKELYDPSAQRYSESDGLVFPDAKIISRTPNGVAIPHGDHYHFIPYSK 300
QY 301 LSALEEKIARWPISTGSTVSTNAKPNEVYSSGLSNPSSLTTSKELSSASDGYIFN 360
Db 301 LSALEEKIARWPISTGSTVSTNAKPNEVYSSGLSNPSSLTTSKELSSASDGYIFN 360
QY 361 PKDIVETATAYIVRGDHFHYIPKSNQIGQPTLPNNSLATPSPSLPIPGTSHKHEED 420
Db 361 PKDIVETATAYIVRGDHFHYIPKSNQIGQPTLPNNSLATPSPSLPIPGTSHKHEED 420
QY 421 GYGFDANRIIADESGFVMSHGDHNYFFKKDLTBEQIKAAQKHLEEVKTSHNGLDLSL 480
Db 421 GYGFDANRIIADESGFVMSHGDHNYFFKKDLTBEQIKAAQKHLEEVKTSHNGLDLSL 480
QY 481 HEODYPGNAKEMKDLKTEETAGTMKQYGVKRESIVVKNKNAIYIPHGHHADPTD 540
Db 481 HEODYPGNAKEMKDLKTEETAGTMKQYGVKRESIVVKNKNAIYIPHGHHADPTD 540
QY 541 EHKPVGIGHSHSYELFKPEEGVAKKGNKVTGSELTNVNLLKNSTFNNQFTLANQ 600
Db 541 EHKPVGIGHSHSYELFKPEEGVAKKGNKVTGSELTNVNLLKNSTFNNQFTLANQ 600
QY 601 KRVSFSPPELEKLGINMLVKLITPDGKVLKSVKQVGEVGNFANFELDPQPLPGOT 660
Db 601 KRVSFSPPELEKLGINMLVKLITPDGKVLKSVKQVGEVGNFANFELDPQPLPGOT 660
QY 661 FKYTIAKDYPEVSYDGTFTVPTSLAYKNASQTIIFYPFHAGDTYLRVNPQFAVPKGTDAL 720
Db 661 FKYTIAKDYPEVSYDGTFTVPTSLAYKNASQTIIFYPFHAGDTYLRVNPQFAVPKGTDAL 720
QY 721 VRVDFEFGHAYLNNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 780
Db 721 VRVDFEFGHAYLNNYKVGELKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 780
QY 781 VPILERENOTDKPSILPQFRKKAQNSKLDKVEPKTSEKVEKLSSETGNSNSTL 840
Db 781 VPILERENOTDKPSILPQFRKKAQNSKLDKVEPKTSEKVEKLSSETGNSNSTL 840
QY 841 EYVPTVDPVQEVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQGN 900
Db 841 EYVPTVDPVQEVAKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEAPQGN 900
QY 901 GENKPSGKSVGTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPSGNGSDPM 960
Db 901 GENKPSGKSVGTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPSGNGSDPM 960
QY 961 LDPALBEEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1019
Db 961 LDPALBEEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1019
```

RESULT 3

AAU84021
ID AAU84021 standard; Peptide; 1019 AA.

XX AAU84021;

DT 08-MAY-2002 (first entry)

DE Truncated variant of *S. pneumoniae* BVH-3, BVH-3M.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW pneumonia; streptococcal bacterial infection; mutant; mutein.

XX
OS
OS
XX
PN
XX
PD
XX
PF
XX
PR
XX
PA
XX
PI
XX
XX
DR
XX
PT
PT
PT
PT
XX
PS
XX

Streptococcus pneumoniae.
Synthetic.

WO200198334-A2.

27-DEC-2001.

19-JUN-2001; 2001WO-CA00908.

20-JUN-2000; 2000US-212683P.

(SHIR-) SHIRE BIOCHEM INC.

Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

WPI; 2002-122272/16.

New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
epitope-bearing polypeptides, useful as vaccine components for treating
or preventing streptococcal infections such as otitis media,
meningitis, and bacteraemia -

Example 1; Page -; 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90%
identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
comprising (I) is useful for therapeutic or prophylactic treatment of
meningitis, otitis media, bacteraemia or pneumonia infection in an
individual susceptible to these disorders. (II) is also useful for
therapeutic or prophylactic treatment of any streptococcal bacterial
infection (e.g., caused by Streptococcus pneumoniae, group A
Streptococcus such as Streptococcus pyogenes, group B Streptococcus or
as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
Staphylococcus aureus) in an individual susceptible to the infection.
A polynucleotide (III) encoding (I) is useful in DNA immunisation
techniques. The Streptococcus polypeptides are useful in a diagnostic
test for *S. pneumoniae* infection. (III) is useful for designing DNA
probes for use in detecting the presence of Streptococcus in a biological
sample suspected of containing the bacteria. The DNA probes may also be
used for detecting circulating *S. pneumoniae* nucleic acid in a sample for
diagnosing streptococcal infections. This sequence represents a truncate
of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
described in the method of the invention.

Note: This sequence does not appear in the specification but has
been created according to information given in the invention.

Sequence 1019 AA:

Query Match 100.0%; Score 5315; DB 23; Length 1019;
Best Local Similarity 100.0%; Pred. No 1.3e-314;
Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAYALNQHRSGENKDNRRVSYVDGSSQSKSENLTDPQVSKQEGIQARQIVIKITDQGYV 60

Db 1 CAYALNQHRSGENKDNRRVSYVDGSSQSKSENLTDPQVSKQEGIQARQIVIKITDQGYV 60

QY 61 TSHGDHYHYNGKVPYDALFSEELLKMDPNYQLKDADIVNEVGKGIKVDGKYYVLKD 120

Db 61 TSHGDHYHYNGKVPYDALFSEELLKMDPNYQLKDADIVNEVGKGIKVDGKYYVLKD 120

QY 121 AAHADNVRTKDEINRQKQEHVKDKNEKNSVAVARSGQRYTTNDGYVFNPAIDIEDTGN 180

Db 121 AAHADNVRTKDEINRQKQEHVKDKNEKNSVAVARSGQRYTTNDGYVFNPAIDIEDTGN 180

QY 181 YIVPHGCHYHYIPKSDLSASELAHAKAHLAGKNQPSQLSYSTASDNTQSVAKGSTSK 240

Db 181 YIVPHGCHYHYIPKSDLSASELAHAKAHLAGKNQPSQLSYSTASDNTQSVAKGSTSK 240

QY 241 PANKSENLOSLKELYDPSAQRYSESDGLVFPDAKIISRTPNGVAIPHGDHYHFIPYSK 300

Db 241 PANKSENLOSLKELYDPSAQRYSESDGLVFPDAKIISRTPNGVAIPHGDHYHFIPYSK 300

```

QY 301 LSALEEKIARVPIISGTSSTVSTAKNEVSVSSLSGSSNPSSLTTSKELSSASDGIYFN 360
Db 301 LSALEEKIARVPIISGTSSTVSTAKNEVSVSSLSGSSNPSSLTTSKELSSASDGIYFN 360
QY 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPSLPINPGTSHKHEED 420
Db 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPSLPINPGTSHKHEED 420
QY 421 GYGFDANRIIAEDSGFVMSHGDNHNFYFKDLTEBOIKAAQKHLEEVKTSNGLDLSLS 480
Db 421 GYGFDANRIIAEDSGFVMSHGDNHNFYFKDLTEBOIKAAQKHLEEVKTSNGLDLSLS 480
QY 481 HEQDYPGNAKEMKLDKKIEEKIAGIMKQYGVKRESIVVANKENALIIYPHGDHHDADPID 540
Db 481 HEQDYPGNAKEMKLDKKIEEKIAGIMKQYGVKRESIVVANKENALIIYPHGDHHDADPID 540
QY 541 EHKPVGIGHSNHYELFKPEGVAKKKGKQVYGEELTNVNLKSTFNQNFYLANGQ 600
Db 541 EHKPVGIGHSNHYELFKPEGVAKKKGKQVYGEELTNVNLKSTFNQNFYLANGQ 600
QY 601 KRVSFPPELEKKGINMLVKLITPDGKVLKYGKVGEGVGNIANFELDQYLPQGT 660
Db 601 KRVSFPPELEKKGINMLVKLITPDGKVLKYGKVGEGVGNIANFELDQYLPQGT 660
QY 661 FYTITASKDYPEVSDGTFTVPTSLAYKMASQITFYFPHAGDTYLRVNPQFVAPKGTDL 720
Db 661 FYTITASKDYPEVSDGTFTVPTSLAYKMASQITFYFPHAGDTYLRVNPQFVAPKGTDL 720
QY 721 VRVDFEFGNAYLENNYKVGKIKPIPKLNOGTRTAGNKIPVTFMANAYLDNQSYIYE 780
Db 721 VRVDFEFGNAYLENNYKVGKIKPIPKLNOGTRTAGNKIPVTFMANAYLDNQSYIYE 780
QY 781 VPILKENQTKPSLLFOFKRKAQENSKLDEKVEEPTSEKVEKSKLSTGNTSNTL 840
Db 781 VPILKENQTKPSLLFOFKRKAQENSKLDEKVEEPTSEKVEKSKLSTGNTSNTL 840
QY 841 BEVPTVDVQEKVAKFAESYGMKLENVLFNNMDGTIELYLPSEGEVKKNNADFTGEAPOGN 900
Db 841 BEVPTVDVQEKVAKFAESYGMKLENVLFNNMDGTIELYLPSEGEVKKNNADFTGEAPOGN 900
QY 901 GENKSENGKYSTGTENQNTENPADSLPEAPNEKPVKPENSTDNQMLNPEGVSGDPM 960
Db 901 GENKSENGKYSTGTENQNTENPADSLPEAPNEKPVKPENSTDNQMLNPEGVSGDPM 960
QY 961 LDPALAEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVKKNLSDFIA 1019
Db 961 LDPALAEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVKKNLSDFIA 1019

RESULT 4
AAB12715
ID AAB12715 standard; Protein: 1039 AA.
XX
AC AAB12715;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-3 protein antigen SEQ ID NO:2.
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
PD
PF 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.

```

The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteraemia and/or pneumonia. The present sequence represents the S. pneumoniae BVH-3 protein antigen.

Query Match 100.0%; Score 5315; DB 21; Length 1039;
 Best Local Similarity 100.0%; Pred. No. 1.3e-314;
 Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CAYALNHRSEKNDNRVSYVDGSSQSKSENITPDQVSQREGIOAEQIVIKITDQGV 60
 21 CAYALNHRSEKNDNRVSYVDGSSQSKSENITPDQVSQREGIOAEQIVIKITDQGV 80
 61 TSHGDHYHYNGKVPYDALFSEELLKMDPNYQKADIVNEVKGYIIKVDGKHYVYLD 120
 81 TSHGDHYHYNGKVPYDALFSEELLKMDPNYQKADIVNEVKGYIIKVDGKHYVYLD 140
 121 AAHADNVRTDEINRQKQEHVKDNEKNSVNAVARSOGRTTNDGVFNPADIETDGA 180
 141 AAHADNVRTDEINRQKQEHVKDNEKNSVNAVARSOGRTTNDGVFNPADIETDGA 200
 191 YIVPHGHHYIYIPKSDLSASELAARAHLAGKNMQPSQSYSTASDNNTQSVAGSTSK 240
 201 YIVPHGHHYIYIPKSDLSASELAARAHLAGKNMQPSQSYSTASDNNTQSVAGSTSK 260
 241 PANKSENQLSKLLEYDPSAQRYSSESDGLVFPDAKTIISRTPNGVAIPHGDHYHFIYPSK 300
 261 PANKSENQLSKLLEYDPSAQRYSSESDGLVFPDAKTIISRTPNGVAIPHGDHYHFIYPSK 320
 301 LSALEEKIARVPIISGTSSTVSTAKNEVSVSSLSGSSNPSSLTTSKELSSASDGIYFN 360
 321 LSALEEKIARVPIISGTSSTVSTAKNEVSVSSLSGSSNPSSLTTSKELSSASDGIYFN 380
 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPSLPINPGTSHKHEED 420
 381 PKDIVEETATAYIVRHGDHFIYIPKSNQIQOPLPNNSLATPSPSLPINPGTSHKHEED 440
 421 GYGFDANRIIAEDSGFVMSHGDNHNFYFKDLTEBOIKAAQKHLEEVKTSNGLDLSLS 480
 441 GYGFDANRIIAEDSGFVMSHGDNHNFYFKDLTEBOIKAAQKHLEEVKTSNGLDLSLS 500
 481 HEQDYPGNAKEMKLDKKIEEKIAGIMKQYGVKRESIVVANKENALIIYPHGDHHDADPID 540
 501 HEQDYPGNAKEMKLDKKIEEKIAGIMKQYGVKRESIVVANKENALIIYPHGDHHDADPID 560
 541 EHKPVGIGHSNHYELFKPEGVAKKKGKQVYGEELTNVNLKSTFNQNFYLANGQ 600
 561 EHKPVGIGHSNHYELFKPEGVAKKKGKQVYGEELTNVNLKSTFNQNFYLANGQ 620
 601 KRVSFPPELEKKGINMLVKLITPDGKVLKYGKVGEGVGNIANFELDQYLPQGT 660
 621 KRVSFPPELEKKGINMLVKLITPDGKVLKYGKVGEGVGNIANFELDQYLPQGT 680

QY 661 FKYTISKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDYLRLVNPQFAVPGKGTDAL 720
 Db 681 FKYTISKDYPEVSYDGTFTVPTSLAYKMASQDIFYPFHAGDYLRLVNPQFAVPGKGTDAL 740
 QY 721 VRYVDFHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYE 780
 Db 741 VRYVDFHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYE 800
 QY 781 VPILKENOTDKPSILPQPKRNKAQENSKLDEKVEPDKTSEKVEKLEKSETGNSNSTL 840
 Db 801 VPILKENOTDKPSILPQPKRNKAQENSKLDEKVEPDKTSEKVEKLEKSETGNSNSTL 860
 QY 841 EYVPTVDPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900
 Db 861 EYVPTVDPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
 QY 901 GENKPSGENKSVGTENOPTENKPADSLPEAPNEKPVKPEPSTDNGLNPNPGENVGSDDPM 960
 Db 921 GENKPSGENKSVGTENOPTENKPADSLPEAPNEKPVKPEPSTDNGLNPNPGENVGSDDPM 980
 QY 961 LDPALPEAPVDPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 1019
 Db 981 LDPALPEAPVDPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSEGEVIKKNLSDFIA 1039

RESULT 5

AAU75932
 ID AAU75932 standard; Protein: 1039 AA.

AC AAU75932;

XX 08-MAY-2002 (first entry)

DE Streptococcus pneumoniae BVH-3 protein version #1.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX pneumonia; streptococcal bacterial infection.

XX Streptococcus pneumoniae.

XX WO200198334-A2.

XX 27-DEC-2001.

XX 19-JUN-2001; 2001WO-CA0908.

XX 20-JUN-2000; 2000US-212683P.

XX (SHIR-) SHIRE BIOCHEM INC.

XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

DR WPI: 2002-122272/16.

DR N-PSDB; ABK15101.

XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -

XX Example 1; Fig 6; 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%

CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of

CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)

CC comprising (I) is useful for therapeutic or prophylactic treatment of

CC meningitis, otitis media, bacteraemia or pneumonia infection in an

CC individual susceptible to these disorders. (II) is also useful for

CC therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A

CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such

CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. noxia or

CC Staphylococcus aureus) in an individual susceptible to the infection.

CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This is the amino acid sequence of
 CC Streptococcus pneumoniae protein BVH-3, used to create the antigenic
 CC peptides described in the method of the invention.

XX Sequence 1039 AA;

Query Match 100.0%; Score 5315; DB 23; Length 1039;
 Best Local Similarity 100.0%; Pred. No. 1.3e-314;
 Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAYALNQHRSEKNNRNVYVDSQSSCKSENLTPOOVSOKEGIAQEQIVIKITDQYV 60
 Db 21 CAYALNQHRSEKNNRNVYVDSQSSCKSENLTPOOVSOKEGIAQEQIVIKITDQYV 80
 QY 61 TSHGDHYHYNGKVPYDALFSEELLMDPNQKLDADIVNEVKGYIIVKDGKYYVYLKD 120
 Db 81 TSHGDHYHYNGKVPYDALFSEELLMDPNQKLDADIVNEVKGYIIVKDGKYYVYLKD 140

QY 121 AAHADNVRTKDEINRQKQEHVKDNEKVNNAVARSQGYTTNDGYVFNPAIDIEDTGNA 180
 Db 141 AAHADNVRTKDEINRQKQEHVKDNEKVNNAVARSQGYTTNDGYVFNPAIDIEDTGNA 200

QY 181 YTVPHGGHYHYIPKSDLSASELAAKAHLAKNMQPSOLSYSSTASDNTOSVAKGSTK 240
 Db 201 YTVPHGGHYHYIPKSDLSASELAAKAHLAKNMQPSOLSYSSTASDNTOSVAKGSTK 260

QY 241 PANKSENQSLKELYDPSAORYSESDGLVDPKAKIISRTPNGVAIPHGDHYHFIPYSK 300
 Db 261 PANKSENQSLKELYDPSAORYSESDGLVDPKAKIISRTPNGVAIPHGDHYHFIPYSK 320

QY 301 LSALEEKARWVPISGTGSTVSTNAKPNEVSSLSGSLSSNPSSLTSTKELASADGYIFN 360
 Db 321 LSALEEKARWVPISGTGSTVSTNAKPNEVSSLSGSLSSNPSSLTSTKELASADGYIFN 380

QY 361 PKDIVEETATAYIVRHGDHFFHYIPKSNQIGOPTLPNNSLATPSPSLINPGTSHKHEED 420
 Db 381 PKDIVEETATAYIVRHGDHFFHYIPKSNQIGOPTLPNNSLATPSPSLINPGTSHKHEED 440

QY 421 GYGFDAIRIIAEDSGFYVSHGDHNNHYFFKDLTEEQKAAQKHLKEEVKTSNGLDLS 480
 Db 441 GYGFDAIRIIAEDSGFYVSHGDHNNHYFFKDLTEEQKAAQKHLKEEVKTSNGLDLS 500

QY 481 HEODYPGNAKEMKDLKKIEEKIAGIMQYGVKRESIVVNKEKNALIVPHGDHHDADPID 540
 Db 501 HEODYPGNAKEMKDLKKIEEKIAGIMQYGVKRESIVVNKEKNALIVPHGDHHDADPID 560

QY 541 EHKPVGIGHSHSNYELFKPEEGVAKKEGNKYVTGBELTNVNVNLLKNSTFNQNTFLANGQ 600
 Db 561 EHKPVGIGHSHSNYELFKPEEGVAKKEGNKYVTGBELTNVNVNLLKNSTFNQNTFLANGQ 620

QY 601 KRVSFSFPELEKKLGINMLVKLITPDGKVLKYSKGVGEGVGNIANFELDQPLPQGT 660
 Db 621 KRVSFSFPELEKKLGINMLVKLITPDGKVLKYSKGVGEGVGNIANFELDQPLPQGT 680

QY 661 FKYTISKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDYLRLVNPQFAVPGKGTDAL 720
 Db 681 FKYTISKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDYLRLVNPQFAVPGKGTDAL 740

QY 721 VRYVDFHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYE 780
 Db 741 VRYVDFHGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIYE 800

QY 781 VPILKENOTDKPSILPQPKRNKAQENSKLDEKVEPDKTSEKVEKLEKSETGNSNSTL 840
 Db 801 VPILKENOTDKPSILPQPKRNKAQENSKLDEKVEPDKTSEKVEKLEKSETGNSNSTL 860

QY 841 EYVPTVDPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900

```

Db      861  EEPVTVDPVQEKVAKFAESYGMKLENVLFNNDGTIELYLPSEGEVKKNNMADFTGEAPQGN 920
Qy      901  GENKPSGKGVSTGVENOPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGVGSDPM 960
Db      921  GENKPSGKGVSTGVENOPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGVGSDPM 980
Qy      961  LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVLFNMDGTIELRLPSGEVKKNNLSDFIA 1019
Db      981  LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVLFNMDGTIELRLPSGEVKKNNLSDFIA 1039

RESULT 6
AAB12750
ID      AAB12750 standard; Protein: 1019 AA.
AC      AAB12750;
XX
DT      21-NOV-2000 (first entry)
DE      Streptococcus pneumoniae strain JNR7/87 BVH-3 protein antigen.
XX
KW      Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW      prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW      otitis media; pneumonia; immunisation; bactericidal.
XX
OS      Streptococcus pneumoniae.
XX
PN      WO200039299-A2.
PD      06-JUL-2000.
PF      20-DEC-1999; 99WO-CA01218.
PR      23-DEC-1998; 98US-O113800.
PA      (BIOC-) BIOCHEM PHARMA INC.
PI      Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
DR      WPI: 2000-452397/39.
XX
PT      Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT      otitis media, bacteraemia and/or pneumonia -
XX
PS      Disclosure; Fig 11; 106pp; English.
XX
CC      The present invention describes nucleic acids (I) encoding protein
CC      antigens (II) from Streptococcus pneumoniae. The protein antigens
CC      have bactericidal activity. The nucleic acids, encoding the protein
CC      antigens, may be used for the recombinant production of the proteins
CC      they encode. The protein antigens may then be used as vaccines for the
CC      prevention and treatment of Streptococcal infections in mammals
CC      (especially humans) which result in, e.g. meningitis, otitis media,
CC      bacteraemia and/or pneumonia. The present sequence represents a
CC      S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ      Sequence 1019 AA;

```

```

Query Match          99.7%; Score 5297; DB 21; Length 1019;
Best Local Similarity 99.7%; Pred. No. 1.6e-313;
Matches 1016; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1  CAYALNHRQENKDNRRSVYDGSQSSQKSENLTDPQVSQKEGIAEQIVIKITDQGV 60
Db      1  CAYALNHRQENKDNRRSVYDGSQSSQKSENLTDPQVSQKEGIAEQIVIKITDQGV 60

Qy      61  TSHGDHYHYNGKVPYDALSEELLKRDPNYQLKADIYNEKVGGLIIKVDGKYVYLKD 120
Db      61  TSHGDHYHYNGKVPYDALSEELLKRDPNYQLKADIYNEKVGGLIIKVDGKYVYLKD 120

Qy      121  AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQGRYTTNDGVYVFNPAADIEDTGN 180

```

```

Db      121  AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQGRYTTNDGVYVFNPAADIEDTGN 180
Qy      181  YIVPHGGHYHYIPKSDLSASELAARAKHLAGKNQWQSLSYSTASDNNTQSVAGKSTSK 240
Db      181  YIVPHGGHYHYIPKSDLSASELAARAKHLAGKNQWQSLSYSTASDNNTQSVAGKSTSK 240
Qy      241  PANKSENQLSKLKYDPSAQRYSESGLVFPDPAKIIISRTPNGVAIPGHDHYHETIPYSK 300
Db      241  PANKSENQLSKLKYDPSAQRYSESGLVFPDPAKIIISRTPNGVAIPGHDHYHETIPYSK 300
Qy      301  LSALAEETARWVPISGTGTVSTNAPNEVYVSSGLSSNPPSSLTTSKLSASOGYIFN 360
Db      301  LSALAEETARWVPISGTGTVSTNAPNEVYVSSGLSSNPPSSLTTSKLSASOGYIFN 360
Qy      361  PKDIVEETATAYIVRHGDHFHYIPKSNQIQPTLPNNSLATPSPLIPNGTSHKHEED 420
Db      361  PKDIVEETATAYIVRHGDHFHYIPKSNQIQPTLPNNSLATPSPLIPNGTSHKHEED 420
Qy      421  GYGFANDRIITAEDESGFVMSHGDNHNYFFKKDLTEQIKAAQKHLEEVKTSNGLDSS 480
Db      421  GYGFANDRIITAEDESGFVMSHGDNHNYFFKKDLTEQIKAAQKHLEEVKTSNGLDSS 480
Qy      481  HEQYDPGNAXEMKDKKIKIEKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDHPID 540
Db      481  HEQYDPGNAXEMKDKKIKIEKIAGIMKQYGVKRESIVVKNKNAIYPHGDHHDHPID 540
Qy      541  BHKPVGICHSHSNYELFKPEEGVAKKEGKNVYTGEBLTNVVNLKNSTFNNQNTLANGQ 600
Db      541  BHKPVGICHSHSNYELFKPEEGVAKKEGKNVYTGEBLTNVVNLKNSTFNNQNTLANGQ 600
Qy      601  KRVSFSPPELEKKGIGINMLVLTTPDGKVLKESVKGVEGSGVGNIANFELDQVYLPQGT 660
Db      601  KRVSFSPPELEKKGIGINMLVLTTPDGKVLKESVKGVEGSGVGNIANFELDQVYLPQGT 660
Qy      661  FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTAL 720
Db      661  FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTAL 720
Qy      721  VRVDFEFGNAYLNNYKVGKIKLPIPKLNQGTTRTAGNKIPVTENMANAYLDNQSTYIVE 780
Db      721  VRVDFEFGNAYLNNYKVGKIKLPIPKLNQGTTRTAGNKIPVTENMANAYLDNQSTYIVE 780
Qy      781  VPILKEKQNTDKSILPQFKRNKAQENSKLDEKVESEPKTSEKVEREKISETGNSNSTL 840
Db      781  VPILKEKQNTDKSILPQFKRNKAQENSKLDEKVESEPKTSEKVEREKISETGNSNSTL 840
Qy      841  EEPVTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQGN 900
Db      841  EEPVTVDPVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVKKNNMADFTGEAPQGN 900
Qy      901  GENKPSGKGVSTGVENOPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGVGSDPM 960
Db      901  GENKPSGKGVSTGVENOPTENKPADSLPEAPNEKVPKPNSTNDGMLNPEGVGSDPM 960
Qy      961  LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVLFNMDGTIELRLPSGEVKKNNLSDFIA 1019
Db      961  LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVLFNMDGTIELRLPSGEVKKNNLSDFIA 1019

RESULT 7
AAB12748
ID      AAB12748 standard; Protein: 1019 AA.
AC      AAB12748;
XX
DT      21-NOV-2000 (first entry)
DE      Streptococcus pneumoniae strain WU2 BVH-3 protein antigen.
XX
KW      Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW      prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW      otitis media; pneumonia; immunisation; bactericidal.
XX

```

OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Disclosure; Fig 11; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

Query Match 99.5%; Score 5291; DB 21; Length 1019;
Best Local Similarity 99.6%; Pred. No. 3.7e-313;
Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAYALNQHRSQENKDNRRYVYDGSOSQKSENLPDQVSQREGIAEQIVIKITDQGVV 60
DB 1 CAYALNQHRSQENKDNRRYVYDGSOSQKSENLPDQVSQREGIAEQIVIKITDQGVV 60
QY 61 TSHGDHYHYNGKVPYDALFSELLMKDPNYOLKADIVNEVKGYYIKVDGKYYVYLDK 120
DB 61 TSHGDHYHYNGKVPYDALFSELLMKDPNYOLKADIVNEVKGYYIKVDGKYYVYLDK 120
QY 121 AAHADNVRTKDEINRQKQEHVDNEKVNNAVARSQGRYTTNDGVFNPADIIEDTGN 180
DB 121 AAHADNVRTKDEINRQKQEHVDNEKVNNAVARSQGRYTTNDGVFNPADIIEDTGN 180
QY 181 YIVPHGHYHYIPKSDLSASELAHAKAHLAGNMOPSQLSYSTASDNNQTQVAKGSTSK 240
DB 181 YIVPHGHYHYIPKSDLSASELAHAKAHLAGNMOPSQLSYSTASDNNQTQVAKGSTSK 240
QY 241 PANKSENLSLLKELYDPSAQYSESDGLVDPKAIISRTENGVAIPHGDIHYHPIYSK 300
DB 241 PANKSENLSLLKELYDPSAQYSESDGLVDPKAIISRTENGVAIPHGDIHYHPIYSK 300
QY 301 LSALAEKIARMVPIISGTGTSTVNTAKNPVSWSLGSLSNPSSLTSTKSLSSASDGYIFN 360
DB 301 LSALAEKIARMVPIISGTGTSTVNTAKNPVSWSLGSLSNPSSLTSTKSLSSASDGYIFN 360
QY 361 PKDIVETATAVIRHGDHFHPIKSNQGTQPTLPNNSLATPSLPINPGTSHKHEED 420
DB 361 PKDIVETATAVIRHGDHFHPIKSNQGTQPTLPNNSLATPSLPINPGTSHKHEED 420
QY 421 GYGFDANRIIAEDESQFVMSHGDHNYHFFKDLTTEQIKAAQKHLEEVKTSNGSLSS 480
DB 421 GYGFDANRIIAEDESQFVMSHGDHNYHFFKDLTTEQIKAAQKHLEEVKTSNGSLSS 480
QY 481 HQDYPGNAKMKDLKKTEEKIAGIMKOYGVKRESIVVNKEKNAILIYPHGHHADPID 540
DB 481 HQDYPGNAKMKDLKKTEEKIAGIMKOYGVKRESIVVNKEKNAILIYPHGHHADPID 540

QY 541 EHKPVGIGHSHSNYELPKPEGVAKKEGNKYVTGELTNVYNLLKNSTFNQNTLANGQ 600
DB 541 EHKPVGIGHSHSNYELPKPEGVAKKEGNKYVTGELTNVYNLLKNSTFNQNTLANGQ 600
QY 601 KRVSFSPPELEKELGGINMLVKLITPDGKVLKESGKVGEGVGNIAFELDQYPLPGQT 660
DB 601 KRVSFSPPELEKELGGINMLVKLITPDGKVLKESGKVGEGVGNIAFELDQYPLPGQT 660
QY 661 FKYTIAASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDGYLRVNPQFAVPGKTAL 720
DB 661 FKYTIAASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDGYLRVNPQFAVPGKTAL 720
QY 721 VRVFDEFHGNAYLENNYKVGGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNSTYIVE 780
DB 721 VRVFDEFHGNAYLENNYKVGGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNSTYIVE 780
QY 781 VPILKEKQTKPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLSNTL 840
DB 781 VPILKEKQTKPSILPOFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTNSTLSNTL 840
QY 841 BEVPTVDVQVKVAKFAESYGMKLENVLFNMDGTIELYLPAGEVIKKNMADFTGEAPQGN 900
DB 841 BEVPTVDVQVKVAKFAESYGMKLENVLFNMDGTIELYLPAGEVIKKNMADFTGEAPQGN 900
QY 901 GENKPESENGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDMGLNPEGNVGSDPM 960
DB 901 GENKPESENGKYSTGTVENQPTENKPADSLPEAPNEKPKVPENSTDMGLNPEGNVGSDPM 960
QY 961 LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMGTIELRLPSGGEVIKKNLSDFLA 1019
DB 961 LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMGTIELRLPSGGEVIKKNLSDFLA 1019

RESULT 8
AAB12749
ID AAB12749 standard; Protein; 1019 AA.
XX AAB12749;
XX 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae strain RX1 BVH-3 protein antigen.
XX Streptococcus pneumoniae: BVH-3; BVH-11; BVH-28; antigen: vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
XX Streptococcus pneumoniae.
XX WO200039299-A2.
XX 06-JUL-2000.
XX 20-DEC-1999; 99WO-CA01218.
XX 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX Disclosure; Fig 11; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

Query Match 99.5%; Score 5291; DB 21; Length 1019;
Best Local Similarity 99.5%; Pred. No. 3.7e-313;
Matches 1014; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAYALNQRSQENKNNRYSYVDGSSQSKSENLTDPQVSQKREGIAEQIVIKITDQGV 60
DB 1 CAYALNQRSQENKNNRYSYVDGSSQSKSENLTDPQVSQKREGIAEQIVIKITDQGV 60

QY 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNYQLKADIYNEVKGGYIIKVDGKYVYLKD 120
DB 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNYQLKADIYNEVKGGYIIKVDGKYVYLKD 120

QY 121 AAHADNVRTKDEINRQKHQVNDNEKNSVAVARSQGRYTTNDGYVFNPAIIEDTGN 180
DB 121 AAHADNVRTKDEINRQKHQVNDNEKNSVAVARSQGRYTTNDGYVFNPAIIEDTGN 180

QY 181 YIVPHGHHYHYIPKSDLSASELAHAGKAGKAGKAGKAGKAGKAGKAGKAGKAGK 240
DB 181 YIVPHGHHYHYIPKSDLSASELAHAGKAGKAGKAGKAGKAGKAGKAGKAGKAGK 240

QY 241 PANKSENQLSKELVDSQAQYSESGLVFPDPAKIIISRTPNGVAIPHGDIYHFIPYSK 300
DB 241 PANKSENQLSKELVDSQAQYSESGLVFPDPAKIIISRTPNGVAIPHGDIYHFIPYSK 300

QY 301 LSALKEERIAEIVISGTSTVSTNAKPNVSVSLGSLSNPSSLTTSKELSSASDGIYFN 360
DB 301 LSALKEERIAEIVISGTSTVSTNAKPNVSVSLGSLSNPSSLTTSKELSSASDGIYFN 360

QY 361 PKOIVETATAYIVRGDHPHYIPKSNQICQPTLPNNSLATPSPILNPGTSHEKHEED 420
DB 361 PKOIVETATAYIVRGDHPHYIPKSNQICQPTLPNNSLATPSPILNPGTSHEKHEED 420

QY 421 GYGFDAANRIIADSESGFVMSGHGHNHYFFKKDLTEPQIRAAQKHLEEVKTSNGLDLS 480
DB 421 GYGFDAANRIIADSESGFVMSGHGHNHYFFKKDLTEPQIRAAQKHLEEVKTSNGLDLS 480

QY 481 HEQDYPGNAMKMDLKKIEKTAGIMQYGVRESIVYVNEKNAIYIPHGDIHHDADPID 540
DB 481 HEQDYPGNAMKMDLKKIEKTAGIMQYGVRESIVYVNEKNAIYIPHGDIHHDADPID 540

QY 541 EHKPVGTGHSNHYELFKPEGVAKKGNKRVYTGEEELTNVNLKSTFNQNTLANGQ 600
DB 541 EHKPVGTGHSNHYELFKPEGVAKKGNKRVYTGEEELTNVNLKSTFNQNTLANGQ 600

QY 601 KRVSFSPPELEKKLGINMLVKLITPDGKLVKSVKGVGEGVGNIANFELDPYLPQGT 660
DB 601 KRVSFSPPELEKKLGINMLVKLITPDGKLVKSVKGVGEGVGNIANFELDPYLPQGT 660

QY 661 FKYTIAASKDPEVSDGTFVPTSLAYKMASQITFYFPHAGDYLVRNPOFAPVPGKTAL 720
DB 661 FKYTIAASKDPEVSDGTFVPTSLAYKMASQITFYFPHAGDYLVRNPOFAPVPGKTAL 720

QY 721 VRVDEFHGNAYLNNYKVGIEKLPKLNQGTTRTAGNIPVTFMANAYLDNQSYIVE 780
DB 721 VRVDEFHGNAYLNNYKVGIEKLPKLNQGTTRTAGNIPVTFMANAYLDNQSYIVE 780

QY 781 VPILEKENQTDKPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTL 840
DB 781 VPILEKENQTDKPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKLSGTGNSNSTL 840

QY 841 BEVPTVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGVYIKKNMADFTGEAPQGN 900
DB 841 BEVPTVPQVEKVAFAESYGMKLENVLFNMDGTIELYLPSEGVYIKKNMADFTGEAPQGN 900

QY 901 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPSNENSTDNGLMPEGNVGS 960
DB 901 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPSNENSTDNGLMPEGNVGS 960

DB 901 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPKVPSNENSTDNGLMPEGNVGS 960
QY 961 LDPALAEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSEGVYIKKNLSOFIA 1019
DB 961 LDPALAEAPADVPQVEKLEKFTASYGLGLDSVIFNMDGTIELRLPSEGVYIKKNLSDLIA 1019

RESULT 9
AAB12752
ID AAB12752 standard; Protein; 1019 AA.
XX AAB12752;
DT 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae strain P4241 BVH-3 protein antigen.
DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX Streptococcus pneumoniae.
XX WO200039299-A2.
XX 06-JUL-2000.
XX 20-DEC-1999; 99WO-CA01218.
XX 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI: 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia
XX Disclosure: Fig 11; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents a
XX S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
SQ Sequence 1019 AA;

Query Match 99.5%; Score 5291; DB 21; Length 1019;
Best Local Similarity 99.6%; Pred. No. 3.7e-313;
Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAYALNQRSQENKNNRYSYVDGSSQSKSENLTDPQVSQKREGIAEQIVIKITDQGV 60
DB 1 CAYALNQRSQENKNNRYSYVDGSSQSKSENLTDPQVSQKREGIAEQIVIKITDQGV 60

QY 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNYQLKADIYNEVKGGYIIKVDGKYVYLKD 120
DB 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNYQLKADIYNEVKGGYIIKVDGKYVYLKD 120

QY 121 AAHADNVRTKDEINRQKHQVNDNEKNSVAVARSQGRYTTNDGYVFNPAIIEDTGN 180
DB 121 AAHADNVRTKDEINRQKHQVNDNEKNSVAVARSQGRYTTNDGYVFNPAIIEDTGN 180

QY 181 YIVPHGHHYHYIPKSDLSASELAHAGKAGKAGKAGKAGKAGKAGKAGKAGKAGK 240
DB 181 YIVPHGHHYHYIPKSDLSASELAHAGKAGKAGKAGKAGKAGKAGKAGKAGKAGK 240

```
QY 241 PANKSENLOSILKELYDPSAORYSESGLVFDPAKIIISRTPNGVAIPHGDHYHFIPIYSK 300
DB 241 PANKSENLOSILKELYDPSAORYSESGLVFDPAKIIISRTPNGVAIPHGDHYHFIPIYSK 300
QY 301 LSALAEKIARMVPIISGTGSTVSTNAKPNEVSSLSGSSNPSLLTTSKELSSASDGYIFN 360
DB 301 LSALAEKIARMVPIISGTGSTVSTNAKPNEVSSLSGSSNPSLLTTSKELSSASDGYIFN 360
QY 361 PKDIVEETATAYIVRHGDHFIHYIPKSNQIQOPTLPNNSLATPSPSLPINPGTSHEKHEED 420
DB 361 PKDIVEETATAYIVRHGDHFIHYIPKSNQIQOPTLPNNSLATPSPSLPINPGTSHEKHEED 420
QY 421 GYGFDANRIIAEDSGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLSS 480
DB 421 GYGFDANRIIAEDSGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLSS 480
QY 481 HEQDYPGNAKEMKDLCKKIEEKIAGIMQYGVKRESIVVNVNKEKNALIIYPHGDHHDADPID 540
DB 481 HEQDYPGNAKEMKDLCKKIEEKIAGIMQYGVKRESIVVNVNKEKNALIIYPHGDHHDADPID 540
QY 541 EHKPVGIGHSHSNYELFPKEGVAKKEGKNYVTGELNVVNLKNSTFNQNTFLANGQ 600
DB 541 EHKPVGIGHSHSNYELFPKEGVAKKEGKNYVTGELNVVNLKNSTFNQNTFLANGQ 600
QY 601 KRVSFSPFPELEKKGILNMLVKLIITPDGKVLKSVKGVFGEVGNIANFELDQVLPQGT 660
DB 601 KRVSFSPFPELEKKGILNMLVKLIITPDGKVLKSVKGVFGEVGNIANFELDQVLPQGT 660
QY 661 FRYTTASKDYPVSDGTFTVPTSLAYKMASQTIIFYPFHAGDTYLRVNPQFAVPGKTDAL 720
DB 661 FRYTTASKDYPVSDGTFTVPTSLAYKMASQTIIFYPFHAGDTYLRVNPQFAVPGKTDAL 720
QY 721 VRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNSTYIIVE 780
DB 721 VRVDFEFGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNSTYIIVE 780
QY 781 VPILKENQTDKPSILPOFKRKNKAGENSKLDEKVEEPTSEKVEKEKLSSETGNSNSTL 840
DB 781 VPILKENQTDKPSILPOFKRKNKAGENSKLDEKVEEPTSEKVEKEKLSSETGNSNSTL 840
QY 841 EEPVTVDPQEVKAVFAESYGMKLENVFNMDGTIELYLPSCGVTKKNMADFTGEAPGN 900
DB 841 EEPVTVDPQEVKAVFAESYGMKLENVFNMDGTIELYLPSCGVTKKNMADFTGEAPGN 900
QY 901 GENKPSNGKSVGTGVNQPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVGSDDPM 960
DB 901 GENKPSNGKSVGTGVNQPTENKPADSLPEAPNEKPKVPKPNSTNDGMLNPEGNVGSDDPM 960
QY 961 LDPALAEAPVDPQVEKLEKFTASGLGLDSVIFNMQGTIELRLPSGVEVIRKKNLSDFIA 1019
DB 961 LDPALAEAPVDPQVEKLEKFTASGLGLDSVIFNMQGTIELRLPSGVEVIRKKNLSDLIA 1019

RESULT 10
AAB12753
ID AAB12753 standard; Protein: 1019 AA.
XX
AC AAB12753;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain A66 BVH-3 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
```

```
XX 20-DEC-1999; 99WO-CA01218.
XX 23-DEC-1998; 98US-Q113800.
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX Disclosure; Fig 11: 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents a
XX S. pneumoniae BVH-3 protein antigen, from the present invention.
XX
XX Sequence 1019 AA;
XX
XX Query Match 99.5%; Score 5291; DB 21; Length 1019;
XX Best Local Similarity 99.6%; Pred. No. 3.7e-313;
XX Matches 1015; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CAYALNHRSEKNDNRVSYVDGSSQSSQSENITPDQVSKQEGIAEQIVKITDGYV 60
DB 1 CAYALNHRSEKNDNRVSYVDGSSQSSQSENITPDQVSKQEGIAEQIVKITDGYV 60
QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYYIIVKDGYYVYVVKD 120
DB 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYYIIVKDGYYVYVVKD 120
QY 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQRYTNDGYVFNADIEDTGN 180
DB 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVAVARSQRYTNDGYVFNADIEDTGN 180
QY 181 YIVPHGGHYHYIPKSDLSASALAAKAGKKNMCPQSLSYSTASDNNTQSVAKGSTSK 240
DB 181 YIVPHGGHYHYIPKSDLSASALAAKAGKKNMCPQSLSYSTASDNNTQSVAKGSTSK 240
QY 241 PANKSENLOSILKELYDPSAORYSESGLVFDPAKIIISRTPNGVAIPHGDHYHFIPIYSK 300
DB 241 PANKSENLOSILKELYDPSAORYSESGLVFDPAKIIISRTPNGVAIPHGDHYHFIPIYSK 300
QY 301 LSALAEKIARMVPIISGTGSTVSTNAKPNEVSSLSGSSNPSLLTTSKELSSASDGYIFN 360
DB 301 LSALAEKIARMVPIISGTGSTVSTNAKPNEVSSLSGSSNPSLLTTSKELSSASDGYIFN 360
QY 361 PKDIVEETATAYIVRHGDHFIHYIPKSNQIQOPTLPNNSLATPSPSLPINPGTSHEKHEED 420
DB 361 PKDIVEETATAYIVRHGDHFIHYIPKSNQIQOPTLPNNSLATPSPSLPINPGTSHEKHEED 420
QY 421 GYGFDANRIIAEDSGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLSS 480
DB 421 GYGFDANRIIAEDSGFVMSHGDHNYHFFKDLTEQIKAAQKHLEEVKTSNGLDLSLSS 480
QY 481 HEQDYPGNAKEMKDLCKKIEEKIAGIMQYGVKRESIVVNVNKEKNALIIYPHGDHHDADPID 540
DB 481 HEQDYPGNAKEMKDLCKKIEEKIAGIMQYGVKRESIVVNVNKEKNALIIYPHGDHHDADPID 540
QY 541 EHKPVGIGHSHSNYELFPKEGVAKKEGKNYVTGELNVVNLKNSTFNQNTFLANGQ 600
DB 541 EHKPVGIGHSHSNYELFPKEGVAKKEGKNYVTGELNVVNLKNSTFNQNTFLANGQ 600
QY 601 KRVSFSPFPELEKKGILNMLVKLIITPDGKVLKSVKGVFGEVGNIANFELDQVLPQGT 660
```


Db 601 KRVSFPPPELEKLGINMLVLTITPDGKLEKVGKVGEGVGNIAFELDQYLPQQT 660
Qy 661 FRYTASQDYPEVSDGFTFTVTSLAYKMASOTIYPPHAGDTYLRVNPQFAVPKGTAL 720
Db 661 FRYTASQDYPEVSDGFTFTVTSLAYKMASOTIYPPHAGDTYLRVNPQFAVPKGTAL 720
Qy 721 VRVDFEFGHAYLENNYKVCETKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSYTIVE 780
Db 721 VRVDFEFGHAYLENNYKVCETKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSYTIVE 780
Qy 781 VPILKENGTDKPSLPQKRNKAQENSKLDKVEEPTSEKVEKEKLTSENGSTNSL 840
Db 781 VPILKENGTDKPSLPQKRNKAQENSKLDKVEEPTSEKVEKEKLTSENGSTNSL 840
Qy 841 EEVPTVDVQKFAVFAESYGMKLENLFNMDGTIELVPSGEVKKNAADFTGEAPOGN 900
Db 841 EEVPTVDVQKFAVFAESYGMKLENLFNMDGTIELVPSGEVKKNAADFTGEAPOGN 900
Qy 901 GENKPSGKSTGTVENQPTENKPADSLPEAPNEKPKVPENSTONGMLNPEGNGVSDPM 960
Db 901 GENKPSGKSTGTVENQPTENKPADSLPEAPNEKPKVPENSTONGMLNPEGNGVSDPM 960
Qy 961 LDPALFAVADVPQKLEKFTASTYGLGLDSVLFNMDGTIELRLPSGEVKKNLSDFA 1019
Db 961 LDPALFAVADVPQKLEKFTASTYGLGLDSVLFNMDGTIELRLPSGEVKKNLSDFA 1019

RESULT 11
AAU84056
ID AAU84056 standard; Peptide; 1238 AA.
XX AAU84056;
XX
XX 08-MAY-2002 (first entry)
XX
XX S. pneumoniae derived chimeric peptide, NEW29.
XX
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumonia; streptococcal bacterial infection; mutant; mutein;
XX BVH-11-2.
XX
XX Streptococcus pneumoniae.
XX Synthetic.
XX OS
XX PN WO200198334-A2.
XX PD 27-DEC-2001.
XX PF 19-JUN-2001; 2001WO-CA00908.
XX PR 20-JUN-2000; 2000US-212683P.
XX PA (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX PI
XX DR WPI: 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX PT or preventing streptococcal infections such as otitis media,
XX PT meningitis, and bacteraemia
XX PS
XX Example 1; Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial

CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novaeccia or
CC Staphylococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a chimeric
CC gene created from fragments and variant fragments of Streptococcus
CC pneumoniae genes, described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.
XX
XX Sequence 1238 AA:
Sg Query Match 81.4%; Score 4325; DB 23; Length 1238;
Best Local Similarity 70.7%; Pred. No. 2.6e-254;
Matches 876; Conservative 39; Mismatches 64; Indels 260; Gaps 12;
Qy 25 SQSSQKSENLT--TPDQVSQKEGIAEQIVIKITDQGVTSHTGHDHYHYNGKVPYDALFSE 82
Db 16 SQRHVESDGLIFDPAQITSR-----TARGVAVPHGNHYHF-----IPYEQMSEL 59
Qy 83 EL-----LKKDPNYQKDA-----EHVKD--NEKV-----147
Db 60 EKRIARIPIRYSNRHWVDPSPRPEPSPQTPPEPSPQAPNPQAPNPIDKLVKEA 119
Qy 99 VNEVKGYYIKVDG-KYYVYVKD-----AAHADNVRTKDEI-----133
Db 120 VRKVGCVGVPEENGVSRYIPAKNLSAETAAGIDSKLAKQESLSHKLGAKKTDLPSDDREF 179
Qy 134 -----NRQKQ-----EHVKD--NEKV-----147
Db 180 YNKAYDLLARIHODLLDNKGRQVDFAELDNLLERLKVSSDKVLDVDDILAFAPIRHPE 239
Qy 148 -----NSNVAVARSGRYTTNDGVVFNPAIDIEDTGNAYIVPHGGHYHYTPKS 195
Db 240 RLGENQAQITTDDEIQVAKLAGYTTEDGYIFDPRITSDGDAYVTPHMTSHWIKKD 299
Qy 196 DLSASELAATAKHL-----209
Db 300 SLSEAFRAAAQAYAKKEKGLTPPSTDHDSGNTAKGAEIYNRVKAACKVPLDRMPYNLQ 359
Qy 210 -----209
Db 360 YTVKNGSLIIPHYDHYHNKIFWFDEGLYEPKGYTLEDLATVYVYVEHPNERPHSD 419
Qy 210 -----AGKNMQPSQLSYSTASDNNTQSVAKGSTSKPANKSENLOSLLKELYDPS 260
Db 420 NGFGNASDHVGNPQPSQLSYSTASDNNTQSVAKGSTSKPANKSENLOSLLKELYDPS 479
Qy 261 AQRYSSEDLVDFDPAKIISRTPNVGAIPHGDHYHFIYPSKLSALEEKIARWVPSGTST 320
Db 480 AQRYSSEDLVDFDPAKIISRTPNVGAIPHGDHYHFIYPSKLSALEEKIARWVPSGTST 539
Qy 321 VSTNAKPNEVVSSLSGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHF 380
Db 540 VSTNAKPNEVVSSLSGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHF 599
Qy 381 HYIPKSNQIGQFTLPNNSLATPSPSLINPGTSHKHEEDGYGFDANRIIAEDSGFVMS 440
Db 600 HYIPKSNQIGQFTLPNNSLATPSPSLINPGTSHKHEEDGYGFDANRIIAEDSGFVMS 659
Qy 441 HGDENHFFKKDLTEQIKAAQKHLVEVKVTSNGLDLSLSHEDDPCGNAKEMDLDKIE 500
Db 660 HGDENHFFKKDLTEQIKAAQKHLVEVKVTSNGLDLSLSHEDDPCGNAKEMDLDKIE 719
Qy 501 EKTAGIKMQYGVKRESIVVNKKNATIIYPHGDHHAADPIDEHKPVGIGHSHSNVELEKPE 560
Db 720 EKTAGIKMQYGVKRESIVVNKKNATIIYPHGDHHAADPIDEHKPVGIGHSHSNVELEKPE 779

QY 561 EGVAKKEGKVVYTGELTNVYVLLKSTFNQNTLANGOKRVSFPPPELKKLGINML 620
Db 780 EGVAKKEGKVVYTGELTNVYVLLKSTFNQNTLANGOKRVSFPPPELKKLGINML 839
QY 621 VKLITPDGKLVLEKVGKVFGEVGNANFELDQYLPFGOTFKYTIASKDYPEVSYDGTFT 680
Db 840 VKLITPDGKLVLEKVGKVFGEVGNANFELDQYLPFGOTFKYTIASKDYPEVSYDGTFT 899
QY 581 VPTSLAYKMASQIFIPFPHAGDTYLRVNPQFAYPKGTDALVRVDFDEFGNAYLENNYKVG 740
Db 900 VPTSLAYKMASQIFIPFPHAGDTYLRVNPQFAYPKGTDALVRVDFDEFGNAYLENNYKVG 959
QY 741 EIKLPIKLNQGTTRAGNKIPVTENANAYLDNQSYIIVEVPILEKENQTDKPSILPQEK 800
Db 960 EIKLPIKLNQGTTRAGNKIPVTENANAYLDNQSYIIVEVPILEKENQTDKPSILPQEK 1019
QY 801 RNKAQENSKLDEKVEBPKEKSEKVEKELSETGNSTNSLEEYVTPDVPQEKVAKFAESY 860
Db 1020 RNKAQENSKLDEKVEBPKEKSEKVEKELSETGNSTNSLEEYVTPDVPQEKVAKFAESY 1079
QY 861 GMKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGNGENKPSGKYSTGTVENQP 920
Db 1080 GMKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGNGENKPSGKYSTGTVENQP 1139
QY 921 TENKPADSLPEAPNEKVPKPEPSTNDGMLNPEGNVSGSDPMLDPALEEAFAVDVPQEKLEK 980
Db 1140 TENKPADSLPEAPNEKVPKPEPSTNDGMLNPEGNVSGSDPMLDPALEEAFAVDVPQEKLEK 1199
QY 981 FTASYGLGLDSVIFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGNGENKPSGKYSTGTVENQP 1019
Db 1200 FTASYGLGLDSVIFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGNGENKPSGKYSTGTVENQP 1238

RESULT 12
AAB12721
ID AAB12721 standard; Protein; 840 AA.
XX AC AAB12721;
XX DT 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae SP63 BVH-3 protein antigen SEQ ID NO:16.
XX DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX OS Streptococcus pneumoniae.
XX WO200039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99WO-CA01218.
XX PR 23-DEC-1998; 98US-0113800.
XX PA (BIOC-) BIOCHEM PHARMA INC.
XX PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI: 2000-452397/39.
XX DR N-PSDB; AAA65738.
XX PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX PT otitis media, bacteraemia and/or pneumonia -
XX PS Claim 18; Fig 19; 106pp; English.
XX CC The present invention describes nucleic acids (I) encoding protein
XX CC antigens (II) from Streptococcus pneumoniae. The protein antigens
XX CC have bactericidal activity. The nucleic acids, encoding the protein

CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae SP63 BVH-3 protein antigen.
XX
SQ Sequence 840 AA;

Query Match 80.5%; Score 4276.5; DB 21; Length 840;
Best Local Similarity 82.3%; Pred. No. 1.4e-251;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNQHRSQENKONNRVSYVDGSSQSKSENLPDQVSQKEGIAQCIYIKITDQGIY 60
Db 1 CAYALNQHRSQENKONNRVSYVDGSSQSKSENLPDQVSQKEGIAQCIYIKITDQGIY 60
QY 61 TSHGDHYHYNGKVPYDALFSELLMKOPNYOLKDADIVNEVKGYYIIKVQDGKYYVYLKD 120
Db 61 TSHGDHYHYNGKVPYDALFSELLMKOPNYOLKDADIVNEVKGYYIIKVQDGKYYVYLKD 120
QY 121 AAHADNVRTKDEINQKOEHVKNKDNKVNNAVARSQRYTTNDGYVFNPAIIEETGNA 180
Db 121 AAHADNVRTKDEINQKOEHVKNKDNKVNNAVARSQRYTTNDGYVFNPAIIEETGNA 180
QY 181 YIVPHGGHYHYTPKSDLSASELAAKAHLAKGNMOPSOLESYSTASDNNTOSVANGSTSK 240
Db 181 YIVPHGGHYHYTPKSDLSASELAAKAHLAKGNMOPSOLESYSTASDNNTOSVANGSTSK 240
QY 241 PANKSENQSLKELYDPSAQRYSQGLVFDPAKIIISRTPNGVAIPHGDHYHFIPYSK 300
Db 241 PANKSENQSLKELYDPSAQRYSQGLVFDPAKIIISRTPNGVAIPHGDHYHFIPYSK 300
QY 301 LSALEEKIARWVPISTGTSTVSTNAKPNEVYSSLSGSLSNPSSLTTSKELSSASDGIYFN 360
Db 301 LSALEEKIARWVPISTGTSTVSTNAKPNEVYSSLSGSLSNPSSLTTSKELSSASDGIYFN 360
QY 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLNNSLATPSPSLPINPQTSHEKHEED 420
Db 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLNNSLATPSPSLPINPQTSHEKHEED 420
QY 421 GYGFDPANRIIADESGFVNSHGDHNYFFPKDLTEEQIAAKQKHLEEVKTSHGNDLSLS 480
Db 421 GYGFDPANRIIADESGFVNSHGDHNYFFPKDLTEEQIAAKQKHLEEVKTSHGNDLSLS 480
QY 481 HQQDYPGNAKEMKDLKDKIEEKIAGIMKQYGVKRESIIVNKEKNAIYYPHGDHHDADID 540
Db 481 HQQDYPGNAKEMKDLKDKIEEKIAGIMKQYGVKRESIIVNKEKNAIYYPHGDHHDADID 540
QY 541 EHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGSELNVVNLKNSTFNQNTLANGQ 600
Db 541 EHKPVGIGHSHSNYELFKPEEGVAKKEGKVVYTGSELNVVNLKNSTFNQNTLANGQ 600
QY 601 KRVSFSPPELEKLGINMLVKLITPDGKLVLEKVGKVFGEVGNANFELDQYLPQGT 660
Db 601 KRVSFSPPELEKLGINMLVKLITPDGKLVLEKVGKVFGEVGNANFELDQYLPQGT 660
QY 661 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQIFIPFPHAGDTYLRVNPQFAVPGKTDAL 720
Db 661 FKYTIASKDYPEVSYDGTFTVPTSLAYKMASQIFIPFPHAGDTYLRVNPQFAVPGKTDAL 720
QY 721 VRVDFEFGNAYLENNYKVGKVFGEVGNANFELDQYLPQGT 780
Db 721 VRVDFEFGNAYLENNYKVGKVFGEVGNANFELDQYLPQGT 780
QY 781 VPILKENQTDKPSILPQFKNKAQENSKLDEKVEBPKEKSEKVEKELSETGNSTNSLT 840
Db 781 VPILKENQTDKPSILPQFKNKAQENSKLDEKVEBPKEKSEKVEKELSETGNSTNSLT 840
QY 841 EEPVTPDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGN 900
Db 841 EEPVTPDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGN 900
QY 900 EEPVTPDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGN 960
Db 900 EEPVTPDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPDSGEVVKKNWADFTGEAPOGN 960

QY 901 GENKPSGKSVSTGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPECNVGSDDPM 960
 ID AAU76151
 XX AAU76151 standard; Protein; 840 AA.
 AC AC
 XX AAU76151;
 DT 08-MAY-2002 (first entry)
 XX Streptococcus pneumoniae BVH-3 protein version #2.
 DE Streptococcus pneumoniae BVH-3 protein version #2.
 XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection.
 XX Streptococcus pneumoniae.
 OS Streptococcus pneumoniae.
 XX W0200198334-A2.
 PN 27-DEC-2001.
 PD 19-JUN-2001; 2001WO-CA00908.
 PF 20-JUN-2000; 2000US-212683P.
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA Hamel J, Ouellet C., Charla C, Charland N, Martin D, Brodeur B;
 PI WPI; 2002-122272/16.
 DR N-PSDB; ABK15105.
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX Example 1; Fig 10; 113pp; English.

CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardia or
 CC Staphylococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This is the amino acid sequence of
 CC Streptococcus pneumoniae protein BVH-3, used to create the antigenic
 CC peptides described in the method of the invention.

XX Sequence 840 AA;

Query Match 80.5%; Score 4276.5; DB 23; Length 840;
 Best Local Similarity 82.3%; Pred. No. 1.4e-251;
 Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNQHRSQENKDNRRVSYVDSQSOKSENLTPOVSOKEGIAEQIVIKITDQSYV 60
 DB 1 CAYALNQHRSQENKDNRRVSYVDSQSOKSENLTPOVSOKEGIAEQIVIKITDQSYV 60
 QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADADIVNEVKGYIIVKDGKYYVYVKD 120
 DB 61 TSHGDHYHYNGKVPYDALFSEELMKDPNYQLKADADIVNEVKGYIIVKDGKYYVYVKD 120
 QY 121 AAHADNVTKDEINRQKQEHVKNDSNVAVARSOGRYTTNDGYVFNPAADIIETDGN 180
 DB 121 AAHADNVTKDEINRQKQEHVKNDSNVAVARSOGRYTTNDGYVFNPAADIIETDGN 180
 QY 181 YIVPHGGHYHYIPKSDLSASELAARAKAHLAKGNMOPSOVSSTASDNNTOSVAKGSTSK 240
 DB 181 YIVPHGGHYHYIPKSDLSASELAARAKAHLAKGNMOPSOVSSTASDNNTOSVAKGSTSK 240
 QY 241 PANKSENLOSLLKELYDSPAQRYSSEDLVDPKAIISRTPNGVAIPHGDHYHFIYSK 300
 DB 224 ----- 223
 QY 301 LSALEEKIARWVPISGTGYSTNAKPNEVYSSLGSLSSNPSSLTTSKELSSASDGIFN 360
 DB 224 ----- 223
 QY 361 PKDIVEETATYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINPOTSHKHEED 420
 DB 224 -----TPSPSLPINPOTSHKHEED 243
 QY 421 GYGFDANRIIAEDSGFVMSHGDHNYFFKDLTEQIKAAOKHLEEVKTSNGLDLSLS 480
 DB 244 GYGFDANRIIAEDSGFVMSHGDHNYFFKDLTEQIKAAOKHLEEVKTSNGLDLSLS 303
 QY 481 HEQDYPGNAKEMKDLCKIEEKIAGIMKOYGVKRESIVVNKEKNAIIPHGCHHRAPIID 540
 DB 304 HEQDYPGNAKEMKDLCKIEEKIAGIMKOYGVKRESIVVNKEKNAIIPHGCHHRAPIID 363
 QY 541 EHKPVGIGHSHSNYELFKPEECVAKKEGKVTGEBELNVNVLKNSTFNQNTLANGO 500
 DB 364 EHKPVGIGHSHSNYELFKPEECVAKKEGKVTGEBELNVNVLKNSTFNQNTLANGO 423
 QY 601 KRVSFSPPELEKKGILNVLKLIITPDGKVLKSVGKVGEGVGNIAFELDQVLPQOT 660
 DB 424 KRVSFSPPELEKKGILNVLKLIITPDGKVLKSVGKVGEGVGNIAFELDQVLPQOT 483
 QY 661 FRYTTASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPHAGDTYLRVNPQFAVPGKTDAL 720
 DB 484 FRYTTASKDYPEVSYDGTFTVPTSLAYKMASQTIIFYPHAGDTYLRVNPQFAVPGKTDAL 543
 QY 721 VRVDEFFHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTMANAYLDNQSYIVE 780
 DB 544 VRVDEFFHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTMANAYLDNQSYIVE 603
 QY 781 VPILKENOTDKPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKILSETGNSTNSL 840
 DB 604 VPILKENOTDKPSILPQFKRKAQENSKLDEKVEEPTSEKVEKEKILSETGNSTNSL 663
 QY 841 EEPVTVDPQVKVAKFAESYGNKLVNFMMDGTIELYLPSCGEVKKNAWDETGEAPOGN 900
 DB 664 EEPVTVDPQVKVAKFAESYGNKLVNFMMDGTIELYLPSCGEVKKNAWDETGEAPOGN 723
 QY 901 GENKPSGKSVSTGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPECNVGSDDPM 960
 DB 724 GENKPSGKSVSTGTVENQPTENKPADSLPEAPNEKPKVPKPNSTDNGLNPECNVGSDDPM 783
 QY 961 LDPALAEAPAVDPQVEKLEKFTASYGLDSDVIFNMDGTIELRPLSFEVGIKKNL 1014
 DB 784 LDPALAEAPAVDPQVEKLEKFTASYGLDSDVIFNMDGTIELRPLSFEVGIKKNL 837
 RESULT 13
 AAU76151
 ID AAU76151 standard; Protein; 840 AA.
 XX AAU76151;
 DT 08-MAY-2002 (first entry)
 XX Streptococcus pneumoniae BVH-3 protein version #2.
 DE Streptococcus pneumoniae BVH-3 protein version #2.
 XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection.
 XX Streptococcus pneumoniae.
 OS Streptococcus pneumoniae.
 XX W0200198334-A2.
 PN 27-DEC-2001.
 PD 19-JUN-2001; 2001WO-CA00908.
 PF 20-JUN-2000; 2000US-212683P.
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA Hamel J, Ouellet C., Charla C, Charland N, Martin D, Brodeur B;
 PI WPI; 2002-122272/16.
 DR N-PSDB; ABK15105.
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX Example 1; Fig 10; 113pp; English.

AC AAU84053;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE S. pneumoniae derived chimeric peptide, NEW26.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutain;
 KW BVH-11-2.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 OS WO200198334-A2.
 PN
 XX
 XX 27-DEC-2001.
 XX
 XX 19-JUN-2001; 2001WO-CA00908.
 PF
 XX 20-JUN-2000; 2000US-212683P.
 PR
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 PI WPI; 2002-122272/16.
 DR
 XX
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia -
 XX
 XX Example 1; Page -: 113pp; English.
 PS
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (III) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocardia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX
 SQ Sequence 1378 AA;
 Query Match 80.2%; Score 4263.5; DB 23; Length 1378;
 Best Local Similarity 53.7%; Pred. No. 1.7e-250;
 Matches 879; Conservative 38; Mismatches 60; Indels 403; Gaps 13;
 QY 25 SSSQSKSENL--TPDQVSQKEGQAQCIYKIKTDQGVYVTHSGDHYHYNGKVPYDALFSE 82
 DB 17 SERHVESDGLIFDPAQITSR-----TARGVAVPHGNHYHF----IPYEQMSL 60
 QY 83 E-----LLMKDQNYQ-----LKDA 96
 DB 61 EKRIARIPIRLYRSHHWPDSEQSPQSTPEPSPSLQFAPNPQAPNPIDKLVKEA 120
 QY 97 DIYNEVKGHHIKVDG-KYIVYIKD-----AAHADNVRKTDEI----- 133

DB 121 --VRKVGQGVFEENGVSRYIPAKDLSAETAAGIDSKIAKQESLSHKLGAACKTDLPSDDR 178
 QY 134 -----NRQKQ-----BHVKD--NEKV----- 147
 DB 179 EFYNKAYDLLARIHQDILLDNKGRQVDFEVLNLLERLKDQVSDKVKLVDDILAFAPIRH 238
 QY 148 -----NSNVAVARSQORYTNDGCVFNPAIDIEDTGNAYIVPHGGHYHIP 193
 DB 239 PERLKSPNAQIITYTDBEIQVAKLAGKYTEDGYIFDPDRITSDGDAYVTPHMTSHWIK 298
 QY 194 KSDLSASELAAAKAH----- 208
 DB 299 KDSLSEAEAAQAAYAKEKGLTPPTDHDQSGNTEAKGAEAIYNVKAACKVPLDRMPYN 358
 QY 209 ----- 208
 DB 359 LOYTVKNGSLIIPHYDHYHNKFEWFEDEGLYEAPKGYLEDLLATYKYVVEHNERPH 418
 QY 209 ----- 208
 DB 419 SDNGFGNASDHYRNKKAQODSKPDEDKHEDEVSEPTHPESDEKENHAGLNPSADNLYKPS 478
 QY 209 ----- 208
 DB 479 TDTEETEAEADTTDEAEIPIQVENSVINAKIADAEALKEKVTDPISIRQNAMEITLGLKSS 538
 QY 209 -----LAKKNQPSQLSVSYSTASNTOSVAKGSTS 239
 DB 539 LLLGTKDNNTISAENVDSLLALLKESQPAIQGNMQPSQLSVSYSTASNTOSVAKGSTS 598
 QY 240 KPANKSENLSLKLKYDPSAQRYSSESDGLVDFPAKIISRTPNGVAIPHGDHYHFIYS 299
 DB 599 KPANKSENLSLKLKYDPSAQRYSSESDGLVDFPAKIISRTPNGVAIPHGDHYHFIYS 658
 QY 300 KLSALEEKIARMPVPSGCTSVSTNAKNEVSVSSLSGSSNPSSLTTSKELSSASDGIIF 359
 DB 659 KLSALEEKIARMPVPSGCTSVSTNAKNEVSVSSLSGSSNPSSLTTSKELSSASDGIIF 718
 QY 360 NPKDVEETATAYIVRHGDHFHYIPKSIQIGOPTLPNNSLATPSPSLPINPCTSEKHEE 419
 DB 719 NPKDVEETATAYIVRHGDHFHYIPKSIQIGOPTLPNNSLATPSPSLPINPCTSEKHEE 778
 QY 420 DOYGFDAIRIADDESFGVMSGDHNNHFFKDLTEEQKAAQKHEEVKTSHNGDLSLS 479
 DB 779 DOYGFDAIRIADDESFGVMSGDHNNHFFKDLTEEQKAAQKHEEVKTSHNGDLSLS 838
 QY 480 SHEQDYPGNAKEMKDLDKKIEKIAQIMKQYGVKRESIVWNKEKNAIYPHGDHHDPI 539
 DB 839 SHEQDYPGNAKEMKDLDKKIEKIAQIMKQYGVKRESIVWNKEKNAIYPHGDHHDPI 898
 QY 540 DEHKPVGISHSHSNYELFKPPEGVAKKSGNKVYTGEELTNVYNLLKNSTFNQNFLLANG 599
 DB 899 DEHKPVGISHSHSNYELFKPPEGVAKKSGNKVYTGEELTNVYNLLKNSTFNQNFLLANG 958
 QY 600 QKRVSEFSPPELEKKLGNLMVLKLTIPDGKVLKYSKGVGFGVGNIANFELDQYLPQ 659
 DB 959 QKRVSEFSPPELEKKLGNLMVLKLTIPDGKVLKYSKGVGFGVGNIANFELDQYLPQ 1018
 QY 660 TKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRYNPOFAVEKGTDA 719
 DB 1019 TKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFPFHAGDTYLRYNPOFAVEKGTDA 1078
 QY 720 LYRVDFEFGHGNAYLENNYKVGKIKPIPKLNGGTTTTRAGNKIPVTFMANAYLDNOSTYIV 779
 DB 1079 LYRVDFEFGHGNAYLENNYKVGKIKPIPKLNGGTTTTRAGNKIPVTFMANAYLDNOSTYIV 1138
 QY 780 EYPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKELSETGNSTNST 839
 DB 1139 EYPILEKENQTDKPSILPQFKRNKAQENSKLDEKVEEPTSEKVEKELSETGNSTNST 1198
 QY 840 LEEVPTDVPQVKAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQG 899
 DB 1199 LEEVPTDVPQVKAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQG 1258

QY 900 NGENKPSNGKSVGTGVTENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPGNNGSDP 959
 ID AAU84057
 XX
 Db 1259 NGENKPSNGKSVGTGVTENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPGNNGSDP 1318
 QY 960 MLDPALEEAFAVDVQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFA 1019
 Db 1319 MLDPALEEAFAVDVQVKLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFA 1378

RESULT 15
 AAU84057
 ID AAU84057 standard; Peptide; 1365 AA.
 XX AC AAU84057;
 XX DT 08-MAY-2002 (first entry)
 XX DE S. pneumoniae derived chimeric peptide, NEW30.
 XX KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 XX KW pneumonia; streptococcal bacterial infection; mutant; mutein;
 XX KW BVH-11-2.
 XX OS Streptococcus pneumoniae.
 XX OS Synthetic.
 XX PN WO200198334-A2.
 XX FN 27-DEC-2001.
 XX PD 19-JUN-2001; 2001WO-CA00908.
 XX PF 20-JUN-2000; 2000US-212683P.
 XX PR (SHIR-) SHIRE BIOCHEM INC.
 XX PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX PI WPI; 2002-122272/16.
 XX DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 XX PT epitope-bearing polypeptides, useful as vaccine components for treating
 XX PT or preventing streptococcal infections such as otitis media,
 XX PT meningitis, and bacteraemia
 XX PS Example 1; Page -; 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90% identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by Streptococcus pneumoniae, group A Streptococcus such as Streptococcus pyogenes, group B Streptococcus such as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or Staphylococcus aureus) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The Streptococcus polypeptides are useful in a diagnostic test for S. pneumoniae infection. (III) is useful for designing DNA probes for use in detecting the presence of Streptococcus in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating S. pneumonia nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of Streptococcus pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has been created according to information given in the invention.
 XX Sequence 1365 AA;

Query Match 80.2%; Score 4261.5; DB 23; Length 1365;
 Best Local Similarity 64.3%; Pred. No. 2.2e-250;
 Matches 878; Conservative 39; Mismatches 62; Indels 387; Gaps 13;

QY 25 SSSSSKSENLL--TPQVQSKEGIAEQIVIKITDQGYVTSBGDHYHYNGKVPYDALFSE 82
 Db 16 SORHVESDGLIFDPAQITSR-----TARGVAVPHGNHYHF-----IPYEQMSEL 59
 QY 83 EL-----LMKDPNVQLKDA-----DI 98
 Db 60 EKRIARIPLRYFSNHWVPSRPEPSQPPTPEPSQPAPNPQAPSPIDSKLVEA 119
 QY 99 VNEKGGYIIKVDG-KYYVYLKD-----AAHADNVRTKDEI-----133
 Db 120 VRKVGDVYFEENGVSRYIPAKNLSAETAAGIDSKLAKOESLSHKLGAKTDLPSOREF 179
 QY 134 -----NROKO-----EHVKD--NEKV-----147
 Db 180 YNKAYDLLARIHQDILLDNKGRQVDFEALDNLRLERKLDYSSDKVLDVDDILAFAPIRHE 239
 QY 148 -----NSNVAVARSQGRYTTNDGYVFNPAIDIEDTGNAYIVPHGHGHHYIPKS 195
 Db 240 RLGGKNAQITYTDDETQVAKLAGYTTEDGYTFDPRDITSDGDAYVTPHMTSHWKKD 299
 QY 196 DLSASELAARAAH-----208
 Db 300 SLSEAEARAAQAYAKEKGLTPPSTDHQDSGNTGAKGAEATYNRVKAARKVPLDRMPYNLQ 359
 QY 209 -----208
 Db 360 YTVKNGSLIIPHYDHYHNKFEWFEDEGLYEAPKGYTLEDLLATVYKVEHPNERPHSD 419
 QY 209 -----208
 Db 420 NGFGNASDHVQRNKGQADTNOTEKPEEKPOTEKPEETPREKPKQSEKPEPKTEEP 479
 QY 209 -----LAGK-----212
 Db 480 EESEPESEEPQVETKEVKEKLEAREADLLGKIQDPIIKSNAKETLTGLKNLLFGTDNN 539
 QY 213 -----NMOPQSLSYSTASDNNTOSVAKGSTSKPANKSENLOSLLK 253
 Db 540 TIMAEAKLALLKESKGNMOPQSLSYSTASDNNTOSVAKGSTSKPANKSENLOSLLK 599
 QY 254 ELYDPSAQRYSSESGLVDFPAKIIISRTPNGVAIPHGDHYHIFIPYKLSALEEKIARMP 313
 Db 600 ELYDPSAQRYSSESGLVDFPAKIIISRTPNGVAIPHGDHYHIFIPYKLSALEEKIARMP 659
 QY 314 ISGTGSTVSTNAKPNEVSSLSGLSSNPSSLTSTKLSASDGYIFNPKDIVEETATAYI 373
 Db 660 ISGTGSTVSTNAKPNEVSSLSGLSSNPSSLTSTKLSASDGYIFNPKDIVEETATAYI 719
 QY 374 VRGHGHFHIPKSNQIOGPTLNNLSLATPSPSLPINTPGTSHKHEEDGYGDFANRIAD 433
 Db 720 VRGHGHFHIPKSNQIOGPTLNNLSLATPSPSLPINTPGTSHKHEEDGYGDFANRIAD 779
 QY 434 ESGFVMSHGDHNYHFFKDLTTEQIKAAQKHLEEVYKTSNGLDLSLSSEHQDYPPGNKEMK 493
 Db 780 ESGFVMSHGDHNYHFFKDLTTEQIKAAQKHLEEVYKTSNGLDLSLSSEHQDYPPGNKEMK 839
 QY 494 DLDKTEEKIAGIMKOYGVKRESIVVNKEKNALIIYPHGDHHDADIDEHKPVGIGHSHSN 553
 Db 840 DLDKTEEKIAGIMKOYGVKRESIVVNKEKNALIIYPHGDHHDADIDEHKPVGIGHSHSN 899
 QY 554 YELFPKEGVAKKEGKVVYTGELTNVNNLLKNSTFNQNTFLANGQKRVSPSPPELEK 613
 Db 900 YELFPKEGVAKKEGKVVYTGELTNVNNLLKNSTFNQNTFLANGQKRVSPSPPELEK 959
 QY 614 KLGINMLVKLIITPDGKLVLEKVSQKVGEGVGNIANFELDQPLPGQTKYTIASKDYBEV 673
 Db 960 KLGINMLVKLIITPDGKLVLEKVSQKVGEGVGNIANFELDQPLPGQTKYTIASKDYBEV 1019
 QY 674 SYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHAYL 733

Db 1020 SYDGTFTVPTSLAYKMASQTFYFPHAGDTYLRVNPQFAVPKGTDALVRVDFHGNAYL 1079
QY 734 ENNYKVCIEIKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKP 793
Db 1080 ENNYKVCIEIKLPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKP 1139
QY 794 SILPQFKRKAQENSKLDEKVEEPKTSKVEKEKLSKTSNSTLNSTLEEVPVDPVQEKV 853
Db 1140 SILPQFKRKAQENSKLDEKVEEPKTSKVEKEKLSKTSNSTLNSTLEEVPVDPVQEKV 1199
QY 854 AKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNMADFTGEAPQNGENKPSSENGKSVST 913
Db 1200 AKFAESYGMKLENVLFNMDGTIELYLPDSGEVIRKKNMADFTGEAPQNGENKPSSENGKSVST 1259
QY 914 GTVENQPTENKPADSLPEAPNEKPVKPENSTDNGMLNPEGNVGSDDPMLDPALEEAFAVDP 973
Db 1260 GTVENQPTENKPADSLPEAPNEKPVKPENSTDNGMLNPEGNVGSDDPMLDPALEEAFAVDP 1319
QY 974 VQEKLEKFTASYGLGLDSVIFNMDGTIELRPLPSGEVIRKKNLSDFIA 1019
Db 1320 VQEKLEKFTASYGLGLDSVIFNMDGTIELRPLPSGEVIRKKNLSDFIA 1365

Search completed: May 13, 2003, 13:53:53
Job time : 52.5434 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:27 ; Search time 23.0524 seconds
(without alignments)
4249.498 Million cell updates/sec

Title: US-09-471-255-55

Perfect score: 5315
Sequence: 1 CAVALNQHRSQENKDNRRVS.....IELRLPSGVEIKNLSDFIA 1019

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73: *
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5297	99.7	1039	2 H95115	conserved hypother
2	5291	99.5	1039	2 D97985	hypotherical prote
3	1292.5	24.3	855	2 D98004	histidine Motif-Co
4	1263.5	23.8	853	2 C97985	hypotherical prote
5	1261	23.7	828	2 E98004	hypotherical prote
6	1255	23.6	802	2 C95136	conserved domain p
7	1235.5	23.2	819	2 B95136	conserved domain p
8	1233	23.2	839	2 C95115	conserved hypother
9	899.5	16.9	822	2 T46758	hypotherical 92.4K
10	324.5	6.1	182	2 F97985	hypotherical prote
11	238.5	4.5	2004	2 B95133	immunoglobulin A1
12	228	4.3	2485	1 H71621	serine/threonine-s
13	223	4.2	1963	2 B98002	IgA-specific metal
14	220	4.1	1701	2 A54498	major merozoite su
15	218	4.1	1873	2 T30944	surface protein pr
16	214	4.0	1939	2 T18372	repeat organellar
17	209	3.9	1701	2 A26868	major merozoite su
18	205.5	3.9	1185	2 A42404	collagen adhesin -
19	203.5	3.8	1849	2 C41859	IgA-specific metal
20	199	3.7	1726	1 SAZQOM	major merozoite su
21	196	3.7	1271	2 A45555	glutamate rich pro
22	196	3.7	1726	2 A45948	major merozoite su
23	193	3.6	3724	2 T18427	hypotherical prote
24	192	3.6	5005	2 F82884	hypotherical prote
25	190.5	3.6	4549	2 T20771	hypotherical prote
26	190.5	3.6	4667	2 T20774	hypotherical prote
27	189	3.6	1038	2 H90053	hypotherical prote
28	187.5	3.5	2144	2 A97942	metalloproteinase
29	187	3.5	1080	2 T43164	Tacta protein - Li

ALIGNMENTS

RESULT 1

H95115

C:Species: conserved hypotherical protein Spi004 [imported] - Streptococcus pneumoniae (strain T
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95115
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
nson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95115
A:Molecule type: DNA
A:Status: preliminary

A:Residues: 1-1039 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:gl4972477; GSPDB:GN00164; TIGR:

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SPI004

Query Match 99.7%; Score 5297; DB 2; Length 1039;

Best Local Similarity 99.7%; Pred. No 9e-254; Mismatches 0; Indels 0; Gaps 0;

Matches 1016; Conservative 0; Indels 0; Gaps 0;

QY	1	CAVALNQHRSQENKDNRRVS	YVDGSSSQSKSENLTPDQVSKQEGIAEQIVIKITDQGVV	60
DB	21	CAVALNQHRSQENKDNRRVS	YVDGSSSQSKSENLTPDQVSKQEGIAEQIVIKITDQGVV	80
QY	61	TSHGDHYHYNGKVPYDALFSELLMKDPNYOLKADIVNEVKGKYYIKYVDGKYYVYVKLD	120	
DB	81	TSHGDHYHYNGKVPYDALFSELLMKDPNYOLKADIVNEVKGKYYIKYVDGKYYVYVKLD	140	
QY	121	AAHADNVRTKDEINRQKOEHVKNKYNVAVARSQRYTTNDGVYFNPADIIETGNA	180	
DB	141	AAHADNVRTKDEINRQKOEHVKNKYNVAVARSQRYTTNDGVYFNPADIIETGNA	200	
QY	181	YIYPHGCHYHYIPKSDLSASELAAKAHLAKGNQPSQLSYSTASDNNNTQSVAKGSTK	240	
DB	201	YIYPHGCHYHYIPKSDLSASELAAKAHLAKGNQPSQLSYSTASDNNNTQSVAKGSTK	260	
QY	241	PANKSENQLLKELYDPSAQRYSESDGLVFPDPAKTIISRTPNGVAIPHGDHVFHFTPYSK	300	
DB	261	PANKSENQLLKELYDPSAQRYSESDGLVFPDPAKTIISRTPNGVAIPHGDHVFHFTPYSK	320	
QY	301	LSALEEKIARVPISGTGTVSTNAKPNVSVSLGSLSSNPSSLTTSKELSSASDGYIFN	360	
DB	321	LSALEEKIARVPISGTGTVSTNAKPNVSVSVSLGSLSSNPSSLTTSKELSSASDGYIFN	380	
QY	361	PKDIVEETATAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLPINFPGTSHKHEED	420	

hypotherical prote
1-phosphatidylinos
rib protein - Stre
rmtB protein (impo
SEC16 protein - yea
SPA2 protein - yea
versican precursor
hypotherical prote
epsin-like protein
hypotherical prote
cell surface antig
mtpr protein - mo
hypotherical prote
hypotherical prote
fibrinogen-binding
hypotherical prote

```

Db 381 PKDIVETATAYIVRHGDHFIYIPKSNQIGQPTLPNNSLATPSPLPINPGTSHKHEED 440
QY 421 GYGFANRIIAEDSGFVMSHGDNHFFKDLTBEQIKAAQKHLEEVKTSNGLDLSLSS 480
Db 441 GYGFANRIIAEDSGFVMSHGDNHFFKDLTBEQIKAAQKHLEEVKTSNGLDLSLSS 500
QY 481 HEQDYPGNKEMKDLKLEETAGIMKQYGVKRESIVVKNKNAIIPHGDDHHPID 540
Db 501 HEQDYPGNKEMKDLKLEETAGIMKQYGVKRESIVVKNKNAIIPHGDDHHPID 560
QY 541 EHPVGIHSHSNYELFKPEEGVAKKKEGKVTGEEITNVNLLKNSFTNNGFTLANGQ 600
Db 561 EHPVGIHSHSNYELFKPEEGVAKKKEGKVTGEEITNVNLLKNSFTNNGFTLANGQ 620
QY 601 KRVSFSPPELEKKGILNMLVKLITPDGKVLKESKVGEGVGNIANFELDPYLPQGT 660
Db 621 KRVSFSPPELEKKGILNMLVKLITPDGKVLKESKVGEGVGNIANFELDPYLPQGT 680
QY 661 FKTYIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLVRNPOFAVPKGTDAL 720
Db 681 FKTYIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLVRNPOFAVPKGTDAL 740
QY 721 VRVDFEHGNAYLENNYKVGKELKIPKLNQGTTRTAGNKIPVTFMANAYLONQSTIYVE 780
Db 741 VRVDFEHGNAYLENNYKVGKELKIPKLNQGTTRTAGNKIPVTFMANAYLONQSTIYVE 800
QY 781 VPILKENTDKPSILPOFKRNKAQENSLDKVEPKTSEKVEKELSETGNSNSTL 840
Db 801 VPILKENTDKPSILPOFKRNKAQENSLDKVEPKTSEKVEKELSETGNSNSTL 860
QY 841 EEPVTPDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900
Db 861 EEPVTPDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
QY 901 GENKPSNGKSVTGTVENQPTENKPADSLPEAPNEKPKVPENSTDMNPEGNVGSDDPM 960
Db 921 GENKPSNGKSVTGTVENQPTENKPADSLPEAPNEKPKVPENSTDMNPEGNVGSDDPM 980
QY 961 LDPALAEAPAVDPQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1019
Db 981 LDPALAEAPAVDPQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039

RESULT 2
D97985
hypothetical protein phtE [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: D97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001.
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D97985
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1039 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:g15458515; GSPDB:GN00174
C:Genetics:
A:Gene: phtE

Query Match 99.5%; Score 5291; DB 2: Length 1039;
Best Local Similarity 99.5%; Pred. No. 1.8e-253;
Matches 1014; Conservative 2; Mismatches 23; Indels 0; Gaps 0;

QY 1 CAYALNHRQENKDNRRSVYDGSQSSQKSENLTDPQVSKQEGIAEQIVIKITDQGVY 60
Db 21 CAYALNHRQENKDNRRSVYDGSQSSQKSENLTDPQVSKQEGIAEQIVIKITDQGVY 80
QY 61 TSHGDHYHYNGKVPYDALFSEELLMKDPYQLKDADIVNEVKGYYIKVDGKYVYVKD 120

```

```

Db 81 TSHGDHYHYNGKVPYDALFSEELLMKDPYQLKDADIVNEVKGYYIKVDGKYVYVKD 140
QY 121 AAHADNVKTKDEINRQKQEHVKQNEKYNVAVARSOGRYTTNDGYVFNPAADIEDTGN 180
Db 141 AAHADNVKTKDEINRQKQEHVKQNEKYNVAVARSOGRYTTNDGYVFNPAADIEDTGN 200
QY 181 YIVPHGGHYHYIPKSDLSASELAALAAKHAHLAKGNQPSQSYSTASDNNTQSVAKGSTK 240
Db 201 YIVPHGGHYHYIPKSDLSASELAALAAKHAHLAKGNQPSQSYSTASDNNTQSVAKGSTK 260
QY 241 PANKSENLOSLLKELYDPSAQRYSSEGLVDFPAKIIISRTPNGVAIPHGDHVFPIYSK 300
Db 261 PANKSENLOSLLKELYDPSAQRYSSEGLVDFPAKIIISRTPNGVAIPHGDHVFPIYSK 320
QY 301 LSALEEKIARVPIISGTSTVSTNAKNEVVSLSLSSNPSSLTTSKELSSASDGYIFN 360
Db 321 LSALEEKIARVPIISGTSTVSTNAKNEVVSLSLSSNPSSLTTSKELSSASDGYIFN 380
QY 361 PKDIVETATAYIVRHGDHFIYIPKSNQIGQPTLPNNSLATPSPLPINPGTSHKHEED 420
Db 381 PKDIVETATAYIVRHGDHFIYIPKSNQIGQPTLPNNSLATPSPLPINPGTSHKHEED 440
QY 421 GYGFANRIIAEDSGFVMSHGDNHFFKDLTBEQIKAAQKHLEEVKTSNGLDLSLSS 480
Db 441 GYGFANRIIAEDSGFVMSHGDNHFFKDLTBEQIKAAQKHLEEVKTSNGLDLSLSS 500
QY 481 HEQDYPGNKEMKDLKLEETAGIMKQYGVKRESIVVKNKNAIIPHGDDHHPID 540
Db 501 HEQDYPGNKEMKDLKLEETAGIMKQYGVKRESIVVKNKNAIIPHGDDHHPID 560
QY 541 EHPVGIHSHSNYELFKPEEGVAKKKEGKVTGEEITNVNLLKNSFTNNGFTLANGQ 600
Db 561 EHPVGIHSHSNYELFKPEEGVAKKKEGKVTGEEITNVNLLKNSFTNNGFTLANGQ 620
QY 601 KRVSFSPPELEKKGILNMLVKLITPDGKVLKESKVGEGVGNIANFELDPYLPQGT 660
Db 621 KRVSFSPPELEKKGILNMLVKLITPDGKVLKESKVGEGVGNIANFELDPYLPQGT 680
QY 661 FKTYIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLVRNPOFAVPKGTDAL 720
Db 681 FKTYIASKDYPEVSYDGTFTVPTSLAYKMASQTIFFYPFHAGDTYLVRNPOFAVPKGTDAL 740
QY 721 VRVDFEHGNAYLENNYKVGKELKIPKLNQGTTRTAGNKIPVTFMANAYLONQSTIYVE 780
Db 741 VRVDFEHGNAYLENNYKVGKELKIPKLNQGTTRTAGNKIPVTFMANAYLONQSTIYVE 800
QY 781 VPILKENTDKPSILPOFKRNKAQENSLDKVEPKTSEKVEKELSETGNSNSTL 840
Db 801 VPILKENTDKPSILPOFKRNKAQENSLDKVEPKTSEKVEKELSETGNSNSTL 860
QY 841 EEPVTPDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 900
Db 861 EEPVTPDVPQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQGN 920
QY 901 GENKPSNGKSVTGTVENQPTENKPADSLPEAPNEKPKVPENSTDMNPEGNVGSDDPM 960
Db 921 GENKPSNGKSVTGTVENQPTENKPADSLPEAPNEKPKVPENSTDMNPEGNVGSDDPM 980
QY 961 LDPALAEAPAVDPQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1019
Db 981 LDPALAEAPAVDPQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVIKKNLSDFIA 1039

RESULT 3
D98004
histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.

```


J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98004
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99864.1; PID:g15458682; GSPDB:GNO0174
C:Genetics:
A:Gene: phpA
C:Superfamily: *Streptococcus agalactiae* hypothetical 92.4k protein

Query Match	24.3%	Score 1292.5;	DB 2;	Length 855;
Best Local Similarity	40.9%	Pred. No. 3.6e-36;		
Matches 304;	Conservative	98;	Mismatches 182;	Indels 159;
				Gaps 21;

Query Match 24.3%; Score 1292.5; DB 2; Length 855;
Best Local Similarity 40.9%; Pred. No. 3.6e-56;
Matches 304; Conservative 98; Mismatches 182; Indels 159; Gaps 21;

QY	1	CAVALNROHSQENKDNRRVSIVDGSOSSKSENLTPDOVSQKREGIOAQEIVIKITDOGV	60
Db	31	CAYELGHLHQAVQYKKNRRVSIVDGKQATQKTENLTDEVSKREGINAEQIVIKITDOGV	90
QY	61	TSBGDHYHYNGVVPYDALFSEBLLMKDPNQLKADIVNEVKGYYIIKVDGKYYVYLD	120
Db	91	TSBGDHYHYNGVVPYDAIISEBLLMKDPNQLKDEIIEIKGGVXIKVDGKYYVYLD	150
QY	121	AHADNVYRTKEINROKQEVKONE---KVNSENVAVARSQGRYTTNDGVFNPADIIED	176
Db	151	AHADNVYRTKEEINROKQESHQREGTPRNDCAVALARSQGRYTTDDGVIFNASDIIED	210
QY	177	TGNAYIVPHGGHHYHYTPKSDLSASELAATAKHAHLAGNMQPSQUSYSSTASDN	228
Db	211	TGDAYIVPHGDHYHYIPKNLSASELAATAKFAUGRNLNSRYTRQNSDNTSRTNWVP	270
QY	229	-----NTQSVAKGSTKXPANKSENLOSLLKXELVDSQAQYSESDGLVFDPKAIISR	280
Db	271	SVSNPGTNTNTNNSNTNSQASQSDNDISLLKQLKPLSQRHVESDGLFDPQAITSR	330
QY	281	TPNGVAIPHGDHYHFYPYKLSALEKRIARWVPISGTGSTVTNNAKPN	328
Db	331	TANGVAIPHGDHYHFYPYQLSPLEEKLARIIPLYRSNHWVDSRPEQSPQSTPEPSP	390
QY	329	--EWSVLSGLSNPSSLITSKE--LSASDCGYFN-----PKDIVEERTAT---AYI	373
Db	391	SPQAPNPQPAPSNPIDEKLVKAEAVKVGQGVYFEENGVPYRIPAKDLSAETAAGDSKL	450
QY	374	VRHGDHFHY-----IPKSN-----QIGOPTLPNNSLATPSPSL-----	406
Db	451	AKQESUSHKLGAKKTDLPSSDRFEYNKAVDILLARIHQDILLDNKQOVDFEALDNLERUK	510
QY	407	-----PI-----NPGTSHEKHE-----EDGYGFGAN	427
Db	511	DVSSDKVKLYDDILAFIPRHPERLCKPNAQITYTDEIQVAKLAGKVTYDEGYFDR	570
QY	428	RIIAEDESCFVMSGHGHNHYFFKDDTEQOIKAAQKHLBEVKTSHNGLSOSLSHEQDYPG	487
Db	571	DITSDEGDAXYTPHMTSHWIKKDSLEAFRAAAQAYAKE-----KGLTPPSTDHOD--SG	624
QY	488	NAKEMKOLDK----KIEEKIAGIMQYGVKRESIVVNKEKNAIYPHGDHGHADPIDEH	542
Db	625	NT-EAKGAEAIIYRVRAAKVPDLDRPYNLQ-----YTVVKNGSLIIPHVDHYH-----	673
QY	543	KPYGIGHSHSNYELFPEEGVAKKGNKYVTGBELTNVYNLL-----KNSTENNON	593
Db	674	-----NIKPEWF--DEGL--YEAPKCYSLDLLATKYVYVEHENRPHSDNGNGFNAS	721
QY	594	FTL---ANGQKRVSFSPPEPLEK	613
Db	722	DHYQRNKNQGADTNOTEKPNEEK	744

RESULT 4
C97985
hypothetical protein phtD [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: C97985
E:Roskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burtett, S.; DeHoff, B.S.
r.R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
Y. P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Author: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Cross-references: GB:AE007317; PID:gl5458514; GSPDB:GN00174
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: C97985
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:Cross-references: GB:AE007317; PID:gl5458514; GSPDB:GN00174
C:Genetics:
A:Gene: phtD
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 23.8%; Score 1263.5; DB 2; Length 853;
Best Local Similarity 41.6%; Pred. NO. 9.6e-55;
Matches 304; Conservative 96; Mismatches 182; Indels 149; Gaps 24;

Qy	1	CAYALNHR	S-QENKDNRRVSYVDGSOSSOKSENLPDQVQSKEGTQAEQVLIKITDQGY	59
Db	20	CSYELGHAQ	VKKESNRYSYIDGQAGQAKENLTPEYSKREGINAEQVLIKITDQGY	79
Qy	60	VTSGDH	HYHYNGKVPYDALFSELLMKDNYQLKADI VNEVKGGYIKKVDGKYVYLK	119
Db	80	VTSGDH	HYHYNGKVPYDAIISELLMKDNYQLKSDI VNEIKGGYIKVDGKYVYLK	139
Qy	120	DAAHADNR	KDEINRQKQEHVKD-NEKYNSENVAVARSGRYTTNDGIVYENPADIIEDTC	178
Db	140	DAAHADNR	KKEIKRQKQFERNHNSRADNAVAAARAQGRVYTTDGYIENASDIEDTC	199
Qy	179	NAYIVPHGG	HYHYIPKSDLSASELAAKAHLACKNMOPSO LSYSSSTASDNNTQ- - -SVAK	235
Db	200	DAYIVPHGD	HYHYIPKSDLSASELAAQAATWNGK- - -OGRSPSSSSHNANPAQRLSENH	257
Qy	236	GSTSKPA	- - -NKSENQSLKELYDPSAQRYSESGLVFPDPAKIIISRTNGVAIPHGDIH	292
Db	258	NLTVTPIY	HOQNGENISSLLRELYAKPLSERHVESDGLIEDPAQITSRTANGVAVPHGDH	317
Qy	293	YHFTPYSK	LASALEBKIAKMPISGTGTVSTNAKPN- - - - -EYVSSLSGSL	338
Db	318	YHFIPY	SQLEKLAIRIPLKRYSRNHWPDSRPEQSPQSTPESPSPQAPNFPQAP	377
Qy	339	SNPSSLAT	SKTSE-LSASDGGIEN- - - - -PKDIVEETAT- - -AYIVRHGDHFHY- - -	382
Db	378	SNPIDEL	VAENVRKVGCGYVFEENGVPYIPAKDLSAETAAGIDSKLAKQESLSHKLGA	437
Qy	383	- - -IPKSN	- - - - -QIGQPLPNNSLATPSSL- - - - -	406
Db	438	KKTDLPS	SDREFYNKAYDILLARTHQLLDNKGQVDFEALDNLRLKDVSSDKVKLVD	497
Qy	407	- - -PI- - -	- - -NPGTSEKH- - - - -EDGYGFQDANRIIAEDESGFYM	439
Db	498	ILAFIAP	IRUPERLGKNAQIYTTDDIOVAKIAGYTTEDGYIFOPRDIITSEGDAYIT	557
Qy	440	SHSDHN	HYFFKKDITEBQIKAAQKHLIEEVTSNGLDLSUSSHQDYPGNAKEMKLDK- -	497
Db	558	PMWTHS	WIKKOSLSEAEAAQAYAKE- - - - -KGLTPPSTDQD- - -SGNT-EAKGAEAY	610
Qy	498	- - -KIEK	IAGIKMKQYGVKRESIVWNKEKNAIYIPGCHDHHADPIDEHKPVGTGHSHSNY	554
Db	611	NRVKAARK	VLDRMPYNLQ- - -YTVVEKNGSLIPIHYDHY- - - - -NIXF	652
Qy	555	ELFKPEG	SVAKKGNKYVTGEETNVNLL- - - - -KNSTFNQNFYL- - -ANGQKR	602
Db	653	EWI- - -	DEGL- - -YFAPKGYSLIEDLATVKKYVVEHNPBRPHSDNGFGNASDHVQRKNQAD	708
Qy	603	VSFSEPP	EELEK	613

```
Db 709 TNOTEKNEEK 719
RESULT 5
E98004
Hypothetical protein phtA [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: E98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
r, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E98004
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1828 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:g15458663; GSPDB:GN00174
C:Genetics:
A:Gene: phtA
C:Superfamily: Streptococcus agalactiae hypothetical 92.4k protein
Query Match 23.7%; Score 1261; DB 2: Length 828;
Best Local Similarity 41.4%; Pred. No. 1.2e-54;
Matches 300; Conservative 95; Mismatches 179; Indels 150; Gaps 19;
QY 1 CAYALNHRSOENKNNRVSYDGSOSSQSENLTDPQVSQKEGIAEQIVIKITDQGV 60
Db 32 CSYELGLYQARTVKENNRVSYIDGKATQKTENLTDPDEVSKREGINAEQIVIKITDQGV 91
QY 61 TSHGDHYHYNGKVPYDALFSEELLKDPNYQLKADIVNEVKGGLIKVDGKYYVYVKD 120
Db 92 TSHGDHYHYNGKVPYDALFSEELLKDPNYQLKADIVNEVKGGLIKVDGKYYVYVKD 151
QY 121 AAHADNVRTKEINRQKQEHVKDNE----KVNNSNVAARSQGRYTTNDGYVFNPAIDIED 176
Db 152 AAHADNVRTKEINRQKQEHVKDNE----KVNNSNVAARSQGRYTTNDGYVFNPAIDIED 211
QY 177 TGNAYIVPHGHHYHYTPKSDLSASELAHAKAHLAGKNMQPSQSYSTASDN----- 228
Db 212 TGNAYIVPHGHHYHYTPKSDLSASELAHAKAHLAGKNMQPSQSYSTASDN----- 271
QY 229 -----NTQSAVGKSTSKPANKSENLSLLKELYDSPSAQRYSESGLVDFPAKIISR 280
Db 272 SVSNPCTTNTNNTSNNTNSQASQNDISLLKQLYKLPLSQRHVESDGLVDFPAQITSR 331
QY 281 TPNGVAIPHGDHYHYTPYKSLSALEEKIARWVP----- 313
Db 332 TARGVAVPHGDHYHYTPYKSLSALEEKIARWVP----- 391
QY 314 -----TSGTGSTVSTNAK-----PNEVYSSLSGLSSNPS 342
Db 392 GPQAPNPKLIDSNLSVLQVRKVGEGYVFEKGISRYVFAKDLPSQVTKNLESKLSQK 451
QY 343 S-----LTTSKELSSASQGYIFNPK-DIVETATATAYIVRHG--DHFHYIPK-----SNQI 389
Db 452 SVSHLTAKKNVAPROQEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLERLNDEST 511
QY 390 GQPLPNNLSLATPSP-SLPINPGTSHEKHE-----EDGYGFANDRIIA 431
Db 512 NKEKLVDDLLAFIAPITHPERLGRKPSQIEYTEDEVRIALQADKYTTSQGYIFDEHDIIS 571
QY 432 EDESGFVMSHGDNHFFKFKDLTEEQIKAAQKHLEE-----VKTSHNGLDSLSS 480
Db 572 DEGDAYVTPHMGSHWIGKDSLSDEKKEVAAQAYTKEKGLPPSPDADVKANPTG-DSAAA 630
QY 481 HEQDYPGNAMKEDLDLKEEKTAGINKQGVKRESIVNKEKNALIIYPHGDHHAHP-- 538
Db 631 IYNRVKG-----EKRIPLVPLPYWV--EHTVEYKNGNLII-PHKDHYHNKFA 675
```

```
QY 539 -IDEH---KPVGIG-----HSHSNVELFKPEEGVAKKBEKNKYVTGEELTNVNNLKNSTFN 590
Db 676 WFDDHTYKAPNGYTLIEDLFATIKYVVEHPDERPHSNDG---WGNASDHVLGKKDHSDEP 731
QY 591 NQNE 594
Db 732 KNFN 735
RESULT 6
C95136
Conserved domain protein Spil175 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95136
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
erson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a Virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95136
A:Molecule type: DNA
A>Status: preliminary
A:Residues: 1802 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75284.1; PID:g14972655; GSPDB:GN00164; TIGR:
A:Experimental source: Strain TIGR4
C:Genetics:
A:Gene: Spil175
C:Superfamily: Streptococcus agalactiae hypothetical 92.4k protein
Query Match 23.6%; Score 1255; DB 2: Length 802;
Best Local Similarity 41.3%; Pred. No. 2.3e-54;
Matches 299; Conservative 95; Mismatches 180; Indels 150; Gaps 19;
QY 1 CAYALNHRSOENKNNRVSYDGSOSSQSENLTDPQVSQKEGIAEQIVIKITDQGV 60
Db 6 CSYELGLYQARTVKENNRVSYIDGKATQKTENLTDPDEVSKREGINAEQIVIKITDQGV 65
QY 61 TSHGDHYHYNGKVPYDALFSEELLKDPNYQLKADIVNEVKGGLIKVDGKYYVYVKD 120
Db 66 TSHGDHYHYNGKVPYDALFSEELLKDPNYQLKADIVNEVKGGLIKVDGKYYVYVKD 125
QY 121 AAHADNVRTKEINRQKQEHVKDNE----KVNNSNVAARSQGRYTTNDGYVFNPAIDIED 176
Db 126 AAHADNVRTKEINRQKQEHVKDNE----KVNNSNVAARSQGRYTTNDGYVFNPAIDIED 185
QY 177 TGNAYIVPHGHHYHYTPKSDLSASELAHAKAHLAGKNMQPSQSYSTASDN----- 228
Db 186 TGNAYIVPHGHHYHYTPKSDLSASELAHAKAHLAGKNMQPSQSYSTASDN----- 245
QY 229 -----NTQSAVGKSTSKPANKSENLSLLKELYDSPSAQRYSESGLVDFPAKIISR 280
Db 246 SVSNPCTTNTNNTSNNTNSQASQNDISLLKQLYKLPLSQRHVESDGLVDFPAQITSR 305
QY 281 TPNGVAIPHGDHYHYTPYKSLSALEEKIARWVP----- 313
Db 306 TARGVAVPHGDHYHYTPYKSLSALEEKIARWVP----- 365
QY 314 -----ISGTGSTVSTNAK-----PNEVYSSLSGLSSNPS 342
Db 366 GPQAPNPKLIDSNLSVLQVRKVGEGYVFEKGISRYVFAKDLPSQVTKNLESKLSQK 425
QY 343 S-----LTTSKELSSASQGYIFNPK-DIVETATATAYIVRHG--DHFHYIPK-----SNQI 389
Db 426 SVSHLTAKKNVAPROQEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLERLNDEST 485
QY 390 GQPLPNNLSLATPSP-SLPINPGTSHEKHE-----EDGYGFANDRIIA 431
Db 486 NKEKLVDDLLAFIAPITHPERLGRKPSQIEYTEDEVRIALQADKYTTSQGYIFDEHDIIS 545
QY 432 EDESGFVMSHGDNHFFKFKDLTEEQIKAAQKHLEE-----VKTSHNGLDSLSS 480
```

```
Db 546 DGDGAYVPHMGHSHWICKDLSOKEVAAQAYTKKGIPEPSPDADVKANPTG--DAAA 604
QY 481 HQDYPGNKAKEMKLDKIEKIACIMKQYGVKRESIVVNKEKNALIIYPHGDHHDADP-- 538
Db 605 IYNRVKG-----EKRIPLVRPYWV--EHTVEVRNGNLII--PHKDHVHNKIFA 649
QY 539 -IDER--KPVGIG----HSHSNVELFPEERGVAKEKGVYTGELTNVYNLLKNSTFN 590
Db 650 WDDHTYKAPNGYTLEDLFATIKYVVEHDERPHSNDG---WGNASEHLVKKDHSDEP 705
QY 591 NONF 594
Db 706 KKNF 709

RESULT 7
B951136
conserved domain protein Spil174 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: B951136
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid-
nson, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B951136
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-819 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75283.1; PID:gl4972654; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: Spil174
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 23.2%; Score 1235.5; DB 2; Length 819;
Best Local Similarity 41.0%; Pred. No. 2.28-53;
Matches 297; Conservative 102; Mismatches 183; Indels 143; Gaps 24;

QY 1 CAYALNQHRS-OENKDNRRVSYVDGSSQSKSENLTDPQVSOKEGIOAEQIVIKITDQGY 59
Db 20 CSYELGRYQAGDKKESRVAIDGDAQAKAENLTPEVSKREGINAEQIVIKITDQGY 79
QY 60 VTSBGDHYHYNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYIITKYDGRYVYLYK 119
Db 80 VTSBGDHYHYNGKVPYDAIIESEELMKDPNYQLKDSIDVNEIKGYIVIKYDGRYVYLYK 139
QY 120 DAAHADNVRTKDEINROKQEHVKD-NEKVNNAVARSQGRYTTNDGYVNPADITDGT 178
Db 140 DAAHADNRTKEETKROKQEH-SHHGGGNSDQAVAAQAQGRYTTDGGYIFNASDIETG 199
QY 179 NAIYVPHGHHYHYIPKSDLSASELAHAKLAGKNQPSQLSYSTASDNNTQ----SVAK 235
Db 200 DAIYVPHGHHYHYIPKNSLASASELAHAKLAGKNQPSQLSYSTASDNNTQ----SVAK 257
QY 236 GSTSKPA---NKENLQSLKELYDSPAQRYSSESDGLVDFPAKTIISRTPNGVATPHGH 292
Db 258 NUTVTPTTHQNGENISSLLRELYAKPUSERHVESDGLIFDPAQTTSRTARGVAVPHGNH 317
QY 293 YHFIYQKSALEEKIARMPVITSGTSTVSTNAKEN-----EYVSSLSGSSNPSSL 344
Db 318 YHFIYQKSALEEKIARMPVITSGTSTVSTNAKEN-----EYVSSLSGSSNPSSL 377
QY 345 TTSKE-LSSASGCGYFN-----PKDVEETAT---ATVIRGDIHFY-----IP 384
Db 378 KLVKEAVRVKGDGYFEENGVSRYIPAKDLSAETAAGDSKLAKQESLSHKLGTKKTDLP 437
QY 385 KSN-----QIGOPTLPNNSLATPPSPSL----- 406
```

```
Db 438 SSSREFYNKAYDILLARTHQDLLDNKQRQVDFEALDNLRLERLKVDSKVKLVEDILAFLA 497
QY 407 PI-----NPTCSHEKHE-----EDGYGPDANRIIIDESEGFVMSHGDHN 445
Db 498 PIRHPELGRPNQAIIITDDEIQVAKLAGKYTTEDGYIFDPRDITSDGDAYVTPHWTSH 557
QY 446 HYFFKDLTTEQIAKQAKHLEEVKTSLSHGLDSSHEQDYPGNAKEMKOLDK-----KIE 500
Db 558 HWIKKDSLSAEARAAQAAYAKE-----KGLTPPSTHDQD--SGNT-EAKGAEATYNRVAAA 610
QY 501 EKTAGIMKQGVKRESIVVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNVELFKPE 560
Db 611 KKVLDRMPYNLQ---YTVGVKSGSLIIPHYDHYH-----NIRFEW--D 650
QY 561 EGVAKEGKNKYVTGEELTNVNNL-----KNSTFNQNTL-----ANGKRVSFSEFP 608
Db 651 EGL--YEAPKGYTLEDLATVKYVVEHDERPHSNDGFGNASDHVQRNKGQADTQTEK 708
QY 509 PELEK 613
Db 709 PSEK 713

RESULT 8
G95115
conserved hypothetical protein SpI003 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: G95115
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
nson, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95115
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75120.1; PID:gl4972476; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SpI003
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 23.2%; Score 1233; DB 2; Length 839;
Best Local Similarity 42.8%; Pred. No. 3e-53;
Matches 295; Conservative 90; Mismatches 164; Indels 140; Gaps 25;

QY 1 CAYALNQHRS-OENKDNRRVSYVDGSSQSKSENLTDPQVSOKEGIOAEQIVIKITDQGY 59
Db 20 CSYELGRYQAGDKKESRVAIDGDAQAKAENLTPEVSKREGINAEQIVIKITDQGY 79
QY 60 VTSBGDHYHYNGKVPYDALFSEELMKDPNYQLKADIVNEVKGYIITKYDGRYVYLYK 119
Db 80 VTSBGDHYHYNGKVPYDAIIESEELMKDPNYQLKDSIDVNEIKGYIVIKYDGRYVYLYK 139
QY 120 DAAHADNVRTKDEINROKQEHVKDNEKVNNSN---VAVARSQGRYTTNDGYVNPADITD 176
Db 140 DAAHADNRTKEETKROKQEH-SHHGGGNSDQAVAAQAQGRYTTDGGYIFNASDIETG 198
QY 177 TGNAYIVPHGHHYHYIPKSDLSASELAHAKLAGKNQPSQLSYSTASDNNTQ----SV 233
Db 199 TGDAYIVPHGHHYHYIPKNSLASASELAHAKLAGKNQPSQLSYSTASDNNTQ----SV 256
QY 234 AKGSTSKPA---NKENLQSLKELYDSPAQRYSSESDGLVDFPAKTIISRTPNGVATPHG 290
Db 257 NENUTVTPTTHQNGENISSLLRELYAKPUSERHVESDGLIFDPAQTTSRTARGVAVPHG 316
QY 291 DHHFIYQKSALEEKIARMPVITSGTSTVSTNAKEN-----EYVSSLSG 336
Db 317 NHHFIYQKSALEEKIARMPVITSGTSTVSTNAKEN-----EYVSSLSG 376
```

```
QY 337 LSSNPSSLTSE-LSASDGYIFN-----PKDIVEETAT---AYIVRHGDHFHY- 382
Db 377 APSNPIDEKLVAEVRKVGVDGYFENGVSRYIPAKDLSAETAGTDSKLAKQESLSHL 436
QY 383 -----IPKSN-----QIGQPTLPN-----NSLA-----TPPSL--- 406
Db 437 GAKKTDLPSSDREFYNKAYDILLARIHQDILLDNKGRQVDFEALDNLRLKDVPSKVLV 496
QY 407 -----PI-----NPQTSHEKHE-----EDGYCFANRIIADESGF 437
Db 497 DQILAPLAPIRHFERLPGKNAQITVDDIQVAKLAGRYTTEDGYIFDFRDTISDEGDAY 556
QY 438 VMSHGDHNYHFKDLEBOIRAAQKHLEEVKTSNGLDLSLSSHEQDYPGNAKEMKDKD 497
Db 557 VTPHMTSHWIKDSLSEAEARAAQAYAKE-----KGLTPSTDHQD-SGNT-EAKGAE 609
QY 498 -----KIEKIAIGIMQYGVKRESIVVNKKNALIIYPHGDHHDHAPIDHKKPVGIGHSHS 552
Db 610 IYNRVRAAKVPLDRPNYLQ---YTVKNGSLIIPHYDHYH-----NI 651
QY 553 NYELFPEEGVAKKCKGNKYVTGCELTNVV 581
Db 652 KPEWF--DEGL--YEAPKGYTLEDLATV 676

RESULT 9
T46758
hypothetical 92.4K protein - Streptococcus agalactiae
C:Species: Streptococcus agalactiae
C>Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
C:Accession: T46758
R:Speilsteberg, B.; Rodzinski, E.; Martin, S.; Weber-Haynenann, J.; Schnitzler, N.; Luet
Infect. Immun. 67, 871-878, 1999
A:Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attachment
A:Reference number: 224091; MUID:99115568; PMID:9916102
A:Accession: T46758
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-822 <SPE>
A:Cross-references: EMBL:AF062533; NID:g4249622; PIDN:AADI3797.1; PID:g4249624
A:Experimental source: strain R268
A:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 16.9%; Score 899.5; DB 2; Length 822;
Best Local Similarity 27.0%; Pred. No. 8.6e-37;
Matches 262; Conservative 144; Mismatches 297; Indels 267; Gaps 31;

QY 2 AYALNQRHQENKNNRVSYVDGSSSQSKS--ENLTPQVSOKEGIAEQIVIKITDQGY 59
Db 22 SYVLGKHHMGLATKNQIAYIDDSKGVKAPKNTKMTDQISAEIGISAEQIVVKITDQGY 81
QY 60 VTSHGDHYHYNGKVPDYDALFSEELLMKDPNYQLKDADIVNEVKGYIKYVDGKYVYVLK 119
Db 82 VTSHGDHYHYNGKVPDYDAIIEELLMTDPNHFQKQSOVINEILDGYIKVNGNYVYVLK 141
QY 120 DAAHADNVTKDEINQKQEHVKD-NEKVNVA-----VARSGRYTTNDGVV 167
Db 142 PGSKRNITIKQIAEQVAKGTEKAKEGLAQVHLSEEAANVEAKRQGYTTDDGYI 201
QY 168 FNPADIIETGNAYIVPHGGHYHYIPKSLASALAAKHA-----GKNMPSQLSYST 224
Db 202 FSPDTIIDGLDAYLVPNGNHYHYIPKDLSPSELAAQAQYKQGRGARDS--DYRPT 259
QY 225 ASDNTQSVAKGSKSPAN-----KSNLQSLKEL 255
Db 260 PAPGRKKAIPDVTNPGOGHQPDNGYHAPPRNDASQNKHQDEKFKGTFKELLQOL 319
QY 256 YDPSQAQRYSESDGLVDFPAKIIISRTPNQVAIPHGDHGHFIPIYSLKLSALEEKIARWPIS 315
Db 320 HRLDLKYRVEEDGLIFEPTQVIKSNFAYVYVPHGDHYHIIPRSQLSPLEMELADRYLAG 379
QY 316 GTGSTVSTNAKNEVYSSLGSLSSNPSSLTTSKELS-----SASD 355
```

```
Db 380 QTDNDNS-----GSDHKPS-----DKVHTFLGHRKAYKGLDGPYDTS 423
QY 356 GYFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPN--NSLATPSPSLPDPGTS 413
Db 424 AYVESKESIHSVDKSGVTAKHGDHFHYI-GFGELEQVELDEVANWVKAKQADDELVAALD 482
QY 414 HEKHEEDGYFD---ANRIIADESGFVMSHGDNHYFEKKDLTEBOIRAAQKHLEEVK 469
Db 483 QEQKREKPL-FDTKKVSKVTKKQVOYIMPKDCKDIFYARYQLDLTQIAFAQELMLKD 541
QY 470 TSHNGLDLSLSEHQDYPGNAKEMKDKDKIEEKIAGIMQYGVKRESIVVNKKNALIIYP 529
Db 542 KKHYRYDI-----VDTGTEPLAVDVSSLPNHAGNATYDTCSSFVI-P 583
QY 530 HGDHHDHP---IDEHFPVGVIGHSHSNYLEKPE--EGVAKKSGKNKYVTGCELTNVNLL 584
Db 584 HIDHIVVVPYSLTRNQIATI-----KYVMQHPEVRPDVMSKPGHE-ESGSVIPNVTPLD 637
QY 585 KNSTFNQNETLANGQKRVSFSPPELEKELGINMLVKLITPDCKVLEKVGKVFGEVGV 644
Db 638 KRAGMPNQI-----IHSAEVQKALAEQ---RFAAPDGYIFD----- 672
QY 645 NIANFELDQPYLPQOTFKYTIASKDYPEVSYDGTFTVPTTSLAYKMASQTIYFPHAGDTY 704
Db 673 -----PDVIAKETP-----VWKDGSFSIPRA----- 694
QY 705 LRVPQFAVPGKTDALVRVDFEFGHNAYLENNYKVGSIKLPKLNQGTTRTAGNKIPVT 764
Db 695 -----DGSRLRINKSDL- 707
QY 765 FMANAYLDNOSTYIVVPILEKEN-----QTDKPSILPQKRNKAQNSKLDKEVEPKT 819
Db 708 -----SOAEQOQACELLAKKAGADATDTPKEEQQ--ADKSNEN-----QOPSE 750
QY 820 SEKYEKELSETGNSNSTLEEVPTVD---PVQEKVAKFAESYGMKLENVLFNMDGTI 875
Db 751 ASKEEKE-----SDDFIDSLPDYGLDRAILEDHINLAQKANIDPKYLIFQFEG-V 800
QY 876 ELYLPSGEVI 885
Db 801 QFYKNGELV 810

RESULT 10
F97985
hypothetical protein phtE-truncation [imported] - Streptococcus pneumoniae (strain R6
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: F97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; Dehoff, B.S.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.W.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: F97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99714.1; PID:gl15458517; GSPDB:GN00174
C:Genetics:
A:Gene: phtE-truncation

Query Match 6.1%; Score 324.5; DB 2; Length 182;
Best Local Similarity 41.0%; Pred. No. 2e-09;
Matches 73; Conservative 26; Mismatches 60; Indels 19; Gaps 5;

QY 344 LTTSKELSSADSGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLPN-NSLATP 402
Db 3 VTPNGVSAVDGQVFNPDIVRDTGDAYIVRHGDHYHYIPKSLNNPPSHSNTTEVGSS 62
```

```

Db 444 AVV-----TDXGET--EVQPESPDTVVSDKGEPEQVAPLPEYKGN-----I 482
Qy 740 GEIK--LPIPKL--NOGTRTAGNKRIPYTEMANAYLDNQSYIVVEVPILKEKENOTDKPSIL 796
Db 483 EQVPEPEVETKKEQGEKT--EEVPV-----KPTETPVNPNEGTTGTSI- 527
Qy 797 PQFRKN---KAQENSKLDEKVEBEKTSEK-----VEKEKUSETGNST 835
Db 528 -QEAENPVQAEETGTTNSEKV-SPDTSKNTGEVSSNPDSSTTSVSGESNKPENHDSKNEN 585
Qy 836 SNSLLEEVPTVDPQVEKVAFAESYGNKKLENLVFNMDGTIELYLPSCVEYKKNKADFTGE 895
Db 586 SEKTVEEVP-VNP-----NDGIVE-----GTSNQETE 611
Qy 896 APQNGENKPSNGKV---STGIVENQPTENKP--ADSLPEAPENKPVKPPENS---TDN 946
Db 612 KPVQPAEETQNSKIANENTGEVSNKPSDSKPPVEESNQPEKNGTATKPPNSGNTTSEN 671
Qy 947 GMLNPE---GNVSGDPMLDPALEAPAVDPQVEKLEKFTASYGLGLDSVIFNMDGTIELR 100
Db 672 GQTFEPGNGSTEDVSTESNTSNSNGNEEIKQENE-----LDPQKKVEPEKTELEUR 724
Qy 1004 LPSGEVIKKNLSD 1016
Db 725 -----NVSD 728

RESULT 12
H71621
serine/threonine-specific protein kinase (EC 2.7.1.1.) PF00150c - malaria p
C:Species: Plasmodium falciparum
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-199
C:Accession: H71621
R:Gardner, M.J.; Rettelien, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.
.: Peritea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium fa
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: H71621
A>Status: preliminary; nucleic acid sequence not shown; translation not sh
A:Molecule type: DNA
A:Residues: 1-2485 <GAR>
A:Cross-references: GB:AE001376; GB:AE001362; NID:g3845108; PIDN:AAC71820.
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF00150C
C:Superfamily: malaria parasite serine/threonine-specific protein kinase P
C:Keywords: phosphotransferase
F:2087-2352/Domain: protein kinase homology <KIN>

Query Match 4.3%; Score 228; DB 1; Length 2485;
Best Local Similarity 18.1%; Pred. No. 0.0068;
Matches 208; Conservative 149; Mismatches 392; Indels 400; Gaps

Qy 3 YALNQHSRQENKNNRVSYVDGSGSSOKSNLNPDQVSQKEGIAEQIVIKITDQGYVTS 62
Db 564 YKBRKSNNNNNNNNIS--SSSSSSSKKHVIT--INKK-----ISS 601
Qy 63 GCDHYHYNGKVPYDALFSBELLM----- 86
Db 602 YNIHYKERKQSFKENFLFFKEKILPSKKDTCVNERQKDLPEKSNHEIKCVSFFNTSD 661
Qy 87 -----KDPNYQLKADATVNEKGGYIKRVDGYYVYLKDAAHADNVRTKDEINRQK 137
Db 662 ISSHSYVKKKEPPFALKNNSIRHPKENNIITYSGKSFNHVQOEKTVLLKKKKEIN--- 718
Qy 138 QEHYKDNKYNVAVARSQGRYITNDGYVFNADI-----IEDTGAYIVIPVGGHY 189
Db 719 DKNTFSCLINHNITF-----YTLQGVNKNLNLGIRDSITYKIDKKNMLKECYNGNN 772
Qy 190 HYIEPKDSLASELAARAHLAGKNMOPSQLSYSTASDNNITQSVAGKTSKPAKNSENLO 249

```

Db 773 DSNKKKKKKLSPS-CDIINDNTP-----YESDEKNNNNNKKSMDFNFKKSKSLY 827
QY 250 SLKELYDS--PSAQRYSDDLFPDPAKIIISRTPNGVAIPHGDHYHPIXPX----- 299
Db 826 NNLSSNRDSTVDHMKYSEYI-----NIQRTKIVELSNKRIRNYKLYSMDFPKVS 881
QY 300 -----KLSALEKERTARVPI----- 314
Db 882 LREKYIDNISNMERYTKNEMINEKISKDDIILYPCDKNKLNSMCPVIENNISREE 941
QY 315 SGTGSTVSTNAKPE-----VSSLSGLSNPSSSLTSKELSSAS 354
Db 942 NEKNSVILKKNNENMFCVGRLCHCMKMNQDNIYDQNIKKNEEIIKHDEYISRE 1001
QY 355 DQYIFNPKDIVEETATYIVRHGDHFHYIPKSNQIQGTPLPNSLATSPSLPINPGTSH 414
Db 1002 EKNKYSK-----CIRNFDDYK-----EQV-----LSY 1025
QY 415 ERHEEDGYGFDAIRIATDE-----SGFVMSHGDHNYHFFKDLTEEQIKAQK 463
Db 1026 HTLDEDKKNDNNLIDMNEAIIETVNGVINIILDRKDNNS-----RDMKEMEKEMEK 1082
QY 464 HLEP--VKTSHGLDLSLSEHQDIPGNAKEMK-DLDKXIEEKIAGIMKO----- 509
Db 1083 KMEKEMEKVMEKMEKMEKEVE-----KELKNEMNNRNMNMNMNEKNIYKNNEIY 1137
QY 510 YGVKRESIVVNEKNAIYPHGDHHADEPDEHK-----PVGIGHSHSNY----- 554
Db 1138 VDMDKELEIVNEKK--LITPFYVES-----DVHKNMNSINNNCKDDIINNLIKKEYVDS 1191
QY 555 -----ELFKPEEGYAKKE-----GNKYIT--GEELTNVYVLLKNSTFNNOFT- 595
Db 1192 CLAQKEENIFRPLNKKDKVKYKRNKNIKTIIHNEEMKRIYQITINKNVFYIYENR 1251
QY 596 -----LANGOKRVSFPPPLEKKLGINMLVKLITPDGKRVLEKVSQKVGEGVN--IANP 649
Db 1252 YENFLIN--HUTYNPK-----NDLFL-----SYKSVNNIRNLYIANK 1289
QY 650 ELDQPY-----LPQO---TFKYTIASKDYPE---VSYDGTFTVPTS LAYKMASQIFYPF 698
Db 1290 HINNYYDYNKLYNQNTYTLKYQVANIDNDHCHKGGGLDYINNMSKECKNR----- 1343
QY 699 HAGDTVLRVNOFAPVPGTDLRVFDEHGNAYL-----ENYKVGEEKLPK 748
Db 1344 -KDKTYL--NKIFHYKKKKDARFFINDEIGSDNYDIKKKYSNDENNYKLN----- 1393
QY 749 LNOGTTTGTAGKIPVTFMANVLDNQSTVIVEVPILEKENQTDKPSILPQFKRKAQENS 808
Db 1394 -----KNMSKSNDEMI---PTLNSEHGNFPSCQPNLLEK---ST 1430
QY 809 KLDEKVEEPTSEKVEKELSTGNSNSTLEEVPTVDPQVEKVAFAESYGMKLENVL 868
Db 1431 TLDNLYDSNMDFTTEKYEVNEN-----DLFNTKRWKFNFSKGNLFNKK 1479
QY 869 F----NWDGTILYLPSEGEVVKKNADFTG-----EAPCGNGKNSKNGKVTG 914
Db 1480 FTVSNEDGVGFSEF-----KNMNLFPKLSNLSKLESVKNNSNNCSNKKGGDDNIG 1531
QY 915 TVENQPTEN 923
Db 1532 NMENMTTN 1540
RESULT 13
B98002
IGA-specific metalloendopeptidase (JG 3.4.24.13) [imported] - Streptococcus pneumoniae
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: B98002
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bursgett, S.; DeHoff, B.S.; E
y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnen, S.; M
e, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: B98002
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1963 <RUP>
A:Cross-references: GB:AE007317; PIDN:AAK99846.1; PID:g15458662; GSPDB:GN00174
A:Gene: iga
C:Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 4.2%; Score 223; DB 2; Length 1963;
Best Local Similarity 22.8%; Pred. No. 0.0083;
Matches 191; Conservative 98; Mismatches 264; Indels 292; Gaps 48;

QY 186 GGHYHYIPKSDLSASELAALAAKHAAGKNMOPSLYS---SSTASDNNNTQSVAKGSTSKFA 242
Db 47 GVHYKYVADSELSEEE-----KKQLVYDIPTYVENDDETYLV----- 84
QY 243 NKSENLOSLLKELYDSPAQRYSESQGLVDPKAKIISRTPNGVAIPHGDHYHPIXPX 302
Db 85 -YKLNQNLQALAE---PNTGSKNERQALVAG---ASLAALGILI-----FAVSKKK 128
QY 303 ALEEKIARMVPISTGS--TVSTNAKPNVSVSLGSLSSNPSSSLTSKELSSASDGYIFN 360
Db 129 VKNKTVLHLVLVAGMGNVSVHLENHLL---LNYN-----TDYELTSGEK--LPL 176
QY 361 PKDVEETATAYIVRHGDHFHYIPKSNQIQGTPLPNSLATSPSLPINPGTSHKEED 420
Db 177 PKEISGYTYIGYI--KEGKTTSDFEVSNO-----EKSAATPT-----KQKV 216
QY 421 GYGFDAIRIATDESGFVMSHGDHNYHFFKDLTEEQIKAQKHLVEKVTSHNGDLSLS 480
Db 217 DYNVTPNEV---DRPSIVQAIQGTTPVSSSTKP-TEVOVVEKPTSTELINPRKEEKSSDS 272
QY 481 HEQDPYGNAKEMKDLKIEKTAGIMKOYGVKRESIVVNEKNAIYPHGDHHADEPID 540
Db 273 QEQ-----LAHKNLETKEEKIS-----PKERTGV-----NTLNQD 305
QY 541 E-----HKPVGIGHSHSNYELFKPEEGVAKKEGKGVYTGTEELTNVYVLLKNSTFNNOF 594
Db 306 EVLSGQINKP-----ELLYRETIE---TKIDFOEEI-----QENP 338
QY 595 TLANGQKRVSFSPFPLEKKIGINM-LVKLIT-----PDGKVLKVSQ 636
Db 339 DLAEQTVRV-----KQEGKLKGVKVEIVRIFSVNKEEVSREIVSTSTAPSPRIVEKGT 392
QY 637 K--VFGE-----GV-----GNTANFELDQYLPQGTFTKYTIASKDYPEVSYDGTFTVPT 683
Db 393 KTVQIKQEPQETGVEHKDVQSGAIVEPAI-QPELP-----EAVVSDKQGEVQCP---TLPE 443
QY 684 SLAYKMASQITIFYPFHAGDTYLRVNPQ---FAVPKQTDALRVFDFHGNAYLENNYK 739
Db 444 AVV-----TDKGET--EQVESPDTVVSQDGEQVAPLPEYKGN-----I 482
QY 740 GEIK--LPIPKL--NQGTTRTAGNKIPVTFMANVLDNQSTVIVEVPILEKENQTDKPSIL 796
Db 483 EQVKPPTPVKTEKQGPKEPT--EEVPV-----KPTETPVNPNNEGTEGTISI- 527
QY 797 PQFKRN---KAQENSKLDEKVEEPTSEKVE-----EKKKLSSETGNS 836
Db 528 -QEAENPVQAPAEESTTNSEKVSQDTSSENTGEVSSNPSSDSTTSVGENSKPEHNSKNENS 586
QY 837 NSTLEEVPTVDPQVEKVAFAESYGMKLENVLFNMDQTIIELYLPSEGEVVKKNADFTGEA 896
Db 587 EKTVEEVP-VNP-----NEGTV-----GTSNQETEX 612
QY 897 PQGNGENKPSNGKV---STGTVENQPTENKP---ADSLPEAPNEKPKVPKPNP-----TDNG 947
Db 613 PVQPAETQTQNSGRKIANIENTGEVSNKFSKSKPPVVEESNQPEKNGKATKPNSENSGNTSNG 672

QY 948 MLNPE 952
 Db 673 QTEPE 677

RESULT 14
 A54498
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (is
 C:Species: Plasmodium falciparum
 C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
 C:Accession: A54498
 R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.;
 Mol. Biochem. Parasitol. 27, 291-302, 1988
 A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmodium
 A:Reference number: A54498; MUID:88142999; PMID:2449612
 A:Accession: A54498
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1701 <PEP>
 A:Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
 C:Superfamily: major merozoite surface antigen
 C:Keywords: surface antigen

Query Match 4.1%; Score 220; DB 2; Length 1701;
 Best Local Similarity 20.3%; Pred. No. 0.0094;
 Matches 246; Conservative 171; Mismatches 395; Indels 402; Gaps 65;

QY 3 YALNQRSEN-----KDNRSYVDGSSQSKSENITPPQVSKGIGIAEQIVIKITDQG 58
 Db 519 YTYNEKQRYNNKFFSSNNSVYVNV---QKLKALSYLED-YSLRKG 561
 QY 59 YVTSHGDDHYHYNGKVPYDA---LFESEL-----LMKDPNYQLKDADIVN- 100
 Db 562 ---SEKDFNHYYTLKTGLEADIKLITEEKSSKENILEKFNKGLTHSANASLEVSDIVKL 618
 QY 101 EVKGGVII-KVGG--KYYYLLKDAHADNVRT-----KDEINRQKQ- 138
 Db 619 QVQVLLIKKIEDLRKIEFLKNAQKDSIHVPNIYKPNQKPEPYLIIVLKKEVDKLKEF 678
 QY 139 -----EHWKDNKENVNVA---VARSGRYTNDG-----YVFNPAIIIEDT 177
 Db 679 IPKVDMLKEQAVLSITQPLVASE---TTEDGGHSTHTLSQGETEVTETETET 735
 QY 178 GNAYIVPHGGHYHI-----PKSDLSASLAAKAHLACKNNQPSQLSYSTASDNNTQS 232
 Db 736 -----VGHTTITVITLPPKEESAPKVKVENSIEHK-----SNDNSQA 774
 QY 233 VAKGSTSKPANK-----SENLSLLKELYD-SPSAQRYSSEGLVDFPAK 276
 Db 775 LRTVYLLKLDEFLKSYCHKYYILVNSNMDQKULEVNLTPPEENELKS-----CDPLD 830
 QY 277 IISRTPNGVAIPHG-----DHYHFTPYSK-----LSALEEK-----IARVVPISG 316
 Db 831 LLFNIGNNIPAMVSLYDSMNIDQLHLFFELYOKEMYIYLLHKEENHIKKLEEQKQITG 890
 QY 317 TSTVST-NAKNEVVSSLGSSLPSSLTTSKELSSASDGVIFNPKOIVETATAYIVR 375
 Db 891 TSTSSPGNTVNTAQSATHSNQSQSNASS---TNTQNG-----VAVSSGPAVVEE 940
 QY 376 HGDHFHYIPKSNQI-GOPTLPN-NSLAPSPSLPINPOTSHKEHEGQYGFDAKRIAE 432
 Db 941 SHDPLTVLSINDLKGIVSLNLGNKTKVPNP-LTIST-TEKEKYEN-----ILKN 990
 QY 433 DSGGFVMSHGDNNHVFKKD-----LTEEIKAAQKHEEYKTS-----HN-----CL 475
 Db 991 NDTYF-----NDDIKQFVKSNSKVIITGLTQKNALNDEIKKLKDLTQLSFDLYNNYKLL 1046
 QY 476 DSLSHSEQDYPGNAREMDK---LDKKIEKIAKIMKQCVKRE-SIVYNKEKNALI----- 527
 Db 1047 DRLFNKKELGQDMQIKLTLKLEQLESKLSLNNPHVNLGNFVFFNKKKEAEIAETE 1106
 QY 528 -----YPHGDHHDADIDEHKPVGI---GSHSHSNYELFKPEEGVAKKE 567

Db 1107 NTLENTKILLKHYKGLVKYNGE---SSPLKLTSEVSIQTEDNYANLEKFRALSIDKGL 1163
 QY 568 GNKVVTGEE-----LTNVNLLKNSTFNQNTFLANGOKRVSEFSPPELEKLGIGNM 619
 Db 1164 NDNHLGKKKLSFLSSGLHLHILTELK-EVINKNIT-----GNSPSENNKA--VNE 1211
 QY 620 LVKL---TIPOGKYLEKSGVKVGEVGNIANFELDQPYLFGQTFKYTI-----ASKDYP 671
 Db 1212 ALKSIVENFLPEAKVTVTVP-----PQPDVTPSLSVRVSSGSGSTREET 1256
 QY 672 EVSYDGTITVPTSLAYKVASQTIFFYPFHAGDTYLRVNFQFAPVPGKTD-----ALV 721
 Db 1257 QIPTSG-----SLTELOQVVQNYDEEDDSLVLPFGESEDNDYLDQVVTGEAIS 1310
 QY 722 RVFDEHGNAYLENNYKVGKIKPLPKLNQGTTRTAGKNIP---VTFMANA----- 769
 Db 1311 VTMDNILSG--FENEYDVILK-PL---AGVYRSLKKQIEKNITFNLNIDILNSLK 1363
 QY 770 ---YLD-----NQSTIVEVPILEKENQTDKPSILPQFRNK-----A 804
 Db 1364 KRKYELDVLESDLAQFKHISSEVYIIE-DSFKLLNSEKNTLLSKYIKESVENDIKFA 1422
 QY 805 QENSKLDEKV-----EEPKTSKVEKEKLSGTGNSNSTLEEVPVDPQVEKVAKE 856
 Db 1423 QEGISYVEKVLAKYKDDLESIKKVIKEEKEKFPSPPTTPS-----PAKTEQKKSKF 1477
 QY 857 A-----ESYGMKLENVLFNM-----DGTIELYL 879
 Db 1478 LPFLTNIELTYNLLNKNKIDDDYLINLKAINDCNVEKDEAHVKITKSLDKAIDKIDLF- 1536
 QY 880 PSGEVYKKNMADFTGEAPQNGENKPSNGK-VSTGTGVENOPT-----ENKPADSIPE 931
 Db 1537 -----KNTNDFRAKKLINDTKKMDLGLSTGLVONFPNTIISKLEGEKQDMINI 1589
 QY 932 APNE--KEVPKPS 943
 Db 1590 SOHQCVKKQCPENS 1603

RESULT 15
 T30944
 surface protein precursor - Enterococcus faecalis
 C:Species: Enterococcus faecalis
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T30944
 R:Shankar, V.; Baghdavan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.
 Infect. Immun. 67, 193-200, 1999
 A:Title: Infection-Gerived Enterococcus faecalis strains are enriched in esp, a gene
 A:Reference number: 220943; MUID:99081742; PMID:9864215
 A:Accession: T30944
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1873 <SHA>
 A:Cross-references: EMBL:AF034779; NID:g3873186; PID:g3873187; PIDN:AAD09858.1

Query Match 4.1%; Score 218; DB 2; Length 1873;
 Best Local Similarity 20.8%; Pred. No. 0.014;
 Matches 215; Conservative 122; Mismatches 377; Indels 322; Gaps 49;

QY 71 NGKYPYDALFSEELLKDPNQLKADIVNEVKGYIIVKDKYVYVLKDAHAD-----N 126
 Db 593 KNNPDEAYEYKWKTEPDTSNVTDSTKGIYTVYKIGNRTFDVDVEFAVRAQAMENDATYVP 652
 QY 127 VRTKDEINRQKQEHVKDNEKVNNSVAVARSQRYTNDGY---VFNPADIIETGNAYI 182
 Db 653 ITTTPETTIQSGKPTFDKDPVP-----LANDAFSVLDVYN-----KDFGNASV 695
 QY 183 VPHGHYHYIFPKSDLSASELAAKAHLAKGNQPSQLSYSTASDNNTQSVAKGSTSKPA 242
 Db 696 DANTGIVTFTPAKGGESEPTGTI-----PIKIVY-----QDGSVGTDLAV 738
 QY 243 NKSENLSQLLKELYDPSAQ-----RYSESGQLVDFPAKIISRTPNGVAIIPHGDHYH 294

Db 739 TVSKN-----IYENPGENIPAGYHKVFTTAGEGTSIESCTTVFAVKDGYSLPE----- 786
QY 295 FIPYKLSALEEKIARMVPISTGTSTVNAK-PNEVSSLSLSSNPSSLLTTSKELSSA 353
Db 787 -----DKLPVLKAKDG-----YTDKWPKE-----ATQPIKADDTFVSSA 822
QY 354 S--DGVIIFNPKD-----IVEETAIVYVRHGDHFIYIPKSNQIGQPTL 394
Db 823 TKLDDIENPGONIPAGYHKVFTTAGEGTSIESGTTVFAVKDG-----VSL 868
QY 395 PKNSLATSPSPILPGTISHEKHEEDGYFDANRIIAEDESQFVMS-----HGDH 444
Db 869 PEDKL-----PVLKAKDGYTDKWPKE-----EATQPIKADDTFVSSATKLDIENPGDN 919
QY 445 -----NHYFFKPD-----LTFEQ-----KAAQKHLE----- 466
Db 920 IPAGYHKVFTTAGEGTSIESGTTVFAVKDGVSLPEDKLPVLKAKDGYTDKWPGEATQPI 979
QY 467 -----EVKTSNGLSLSSHEODYPCNAKEMKDLOKIEEKIAGIMKOYGVKRESIVNWK 521
Db 980 KADDTFVSSATKLDKSDADKXTPGQKVTTELNKE-PDASEGINKKDLKPKDAKYTWK 1038
QY 522 E-----KNAIYPHGDDHADIPIDEHKPVGIGHSHSNYELFKPE-EGVAKKEGK 570
Db 1039 EKVDISTAGNKKGTVVVTVSD--GSSDEVE--VDVIVTDNRSDADKYEPTVEGEVEIGK 1095
QY 571 VYTGEBELTNVNLKKNSTFNN-----QNFILANGOKRVSFSPPELEKKGILNMLVKL 623
Db 1096 VDLTDNVTNLPTLPQGTIVTDVTPGGTIDTNTPGNVEGVIEVTPD-----GTRDVKV 1149
QY 624 ---ITPDGKYLEKVSCKVEGV-----GNIANFELDQPYLPQGTIFYTIASKDYP 671
Db 1150 PVEVTDNRSDADKYEPTVEGEKVEIGKVDLTDNVTNL-----PTLPQGT---TVT----- 1197
QY 672 EVSYDGTFTVPTSLAYKMASOTIFYPFHAGDTYLRVNPQFAVPKGTDAIVRVFDEPHGNA 731
Db 1198 DVTPGGTIDTNTPGNVEGVIEV-----TY-----PDGKOTVKVPVEVTDNR 1239
QY 732 YLENNY-----KV---GEIKLP-----IPKLNQGT---RTAGNKIPVTFMANAYLDN 773
Db 1240 SDADKYEPTVEGEKVEIGKVDLTDNVTNLPTLPQGTIVTDVTPGGTIDTNTPGNVEGVI 1299
QY 774 QSTY-----IVEVPILEKENQTDKPSILPOPKRKAQENSKLDEKVESEPKTSEKVEK 827
Db 1300 EVTPDGTCTKTVKVPVEVTDNRSDADKXTPMVEGEKVEIGKVD----- 1343
QY 828 LSETGNSTNSLTLEEVPDPOEKVAKFAESYGMKLENVFNMDGTIELYLPFG--EVI 885
Db 1344 --LTDNVTNLPTLPQGTIVTDVTPG-----GTIDTNTPGNVEGVIEVTPDGTCTKTV 1393
QY 886 K-----KNWADTGTGAPQGNKPSNGKVS-TGTVENQPTENKPADSLPEAPNEKPV 938
Db 1394 KVPVEVTDNRSDADKYEPTVEGE-KVEIGKVDLTDNVTNLPT-----LPQGTIVTDV 1445
QY 939 KPNSTDNGMLNPEGN 954
Db 1446 TPGGTIDT---NTPGN 1458

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:50:17 ; Search time 11.9531 Seconds
(without alignments)
3535.857 Million cell updates/sec

Title: US-09-471-255-55

Perfect score: 5315

Sequence: 1 CAYALNQHRQENKNNRVS.....IELRLPSGEVIKKNLSDFIA 1019

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	221	4.2	1701	1 MSP1_PLAFM	P08569 plasmodium
2	220	4.1	1701	1 MSP1_PLAF	P13819 plasmodium
3	205.5	3.9	1183	1 CRA_STRAU	Q53654 staphylococ
4	203.5	3.8	1849	1 IGA4_HAEIN	P45386 haemophilus
5	199	3.7	1726	1 MSP1_PLAF	P04934 plasmodium
6	199	3.7	1726	1 MSP1_PLAF	P50495 plasmodium
7	195.5	3.7	1637	1 MSP1_PLAF	P80544 staphylococ
8	189.5	3.6	1531	1 Y038_CAEEL	Q09459 caenorhabdi
9	186.5	3.5	1570	1 P3K1_DICDI	P54673 dictyosteli
10	185	3.5	2195	1 SC16_YEAST	P48415 saccharomyc
11	184.5	3.5	1466	1 SPA2_YEAST	P23201 saccharomyc
12	184	3.5	3381	1 PGCV_BOVIN	P81282 bos taurus
13	181.5	3.4	1682	1 MSP1_PLAF3	P19598 plasmodium
14	179.5	3.4	1636	1 BUD3_YEAST	P25558 saccharomyc
15	178.5	3.4	2738	1 PGCV_RAT	Q08764 rattus norv
16	177	3.3	1702	1 IGA2_HAEIN	P45384 haemophilus
17	176.5	3.3	1139	1 HWI_MYCGE	Q49413 mycoplasma
18	176.5	3.3	1658	1 Y067_YEAST	Q03661 saccharomyc
19	175.5	3.3	2748	1 NMU1_YEAST	Q00402 saccharomyc
20	174.5	3.3	1577	1 HLYA_PROMI	P16466 proteus mir
21	174.5	3.3	1807	1 VTA2_XENLA	P18709 xenopus lae
22	174.5	3.3	2485	1 PTND_HUMAN	Q12923 homo sapien
23	173	3.3	1694	1 IGA0_HAEIN	P44969 haemophilus
24	173	3.3	1790	1 US01_YEAST	P25386 saccharomyc
25	172.5	3.2	1142	1 GIN4_YEAST	Q12263 saccharomyc
26	172.5	3.2	1532	1 IGA_NEIGO	P09790 neisseria g
27	172.5	3.2	1781	1 AKAC_HUMAN	Q02952 homo sapien
28	172	3.2	817	1 YG4A_YEAST	P46949 saccharomyc
29	172	3.2	2869	1 RBPL_PLAFB	Q00798 plasmodium
30	171	3.2	710	1 L78_ARAPH	Q06738 arabidopsis
31	169.5	3.2	1167	1 SCAL_STRPY	P15926 streptococ
32	168.5	3.2	818	1 PTK2_YEAST	P47116 saccharomyc
33	168.5	3.2	1435	1 EBAL_PLAF	P19214 plasmodium

ALIGNMENTS

RESULT 1

ID	MSP1_PLAFM	STANDARD	PRT	1701 AA
AC	P08569:			
DT	01-AUG-1998 (Rel. 08, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens)			
DE	(PMW58) (P190).			
GN	MSP-1.			
OS	Plasmodium falciparum (isolate mad20 / Papua New Guinea).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_Taxid=70153;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88011243; PubMed=3079521;			
RA	Tanabe K., Mackay M., Goman M., Scaife J.G.;			
RT	"Allelic dimorphism in a surface antigen gene of the malaria parasite			
RT	Plasmodium falciparum.";			
RL	J. Mol. Biol. 195:273-287(1987).			
RN	[2]			
RP	REVISIONS TO 1403; 1569 AND 1629.			
RA	Tanabe K.;			
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1-115 FROM N.A.			
RX	MEDLINE=86136024; PubMed=3004972;			
RA	Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,			
RT	Stunnenberg H., Buijard H.;			
RT	"Polymorphism of the precursor for the major surface antigens of			
RT	Plasmodium falciparum merozoites: studies at the genetic level.";			
RL	EMBO J. 4:3823-3829(1985).			
CC	-!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor			
CC	(Potential).			
CC	-!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42			
CC	kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF			
CC	MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; X05624; CAA29112.1;			
DR	PIR; A26868; A26868.			
DR	PIR; B25120; B25120.			
DR	InterPro; IPR000561; EGF-like.			
DR	Pfam; PF00008; EGF; 1.			
DR	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;			
KW	Transmembrane; GPI-anchor.			
KW	SIGNAL 1 19 POTENTIAL.			
FT	CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.			
FT	CARBOHYD 110 110 N-LINKED (GLCNAC...) (POTENTIAL).			

P07866 saccharomyc
P04932 plasmodium
Q01484 homo sapien
P04933 plasmodium
Q01550 xenopus lae
P52081 staphylococ
P27951 streptococ
P47035 saccharomyc
Q06105 saccharomyc
Q04958 saccharomyc
P45385 haemophilus
Q09904 schizosacch

FT CARBOHYD 239 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 470 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 536 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 607 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 802 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 899 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 919 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 965 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 991 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 1089 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 1196 N-LINKED (GLCNAC . . .) (POTENTIAL).
 FT CARBOHYD 1588 N-LINKED (GLCNAC . . .) (POTENTIAL).
 SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 4.2%; Score 221; DB 1; Length 1701;
 Best Local Similarity 20.3%; Pred. No. 0.0085;
 Matches 246; Conservative 172; Mismatches 394; Indels 402; Gaps 65;

QY 3 YALNQRSQEN-----KDNRRSVSDGSSQSSQSENLTPOVQSKQEGIQAEQIVIKITDOG 58
 Db 519 YTYNVEKQRTNNRNFSSNNVNV---QKLKALSYLEL-YSLRGI----- 561
 QY 59 YVTSHGHHYHYNGKVPYDA---LFSSEL-----LMKDPNYQLKDADIVN- 100
 Db 562 ---SEKDFNHYHTLGTGLEADIKKLTETEEKSSSENKILEKNFKGLTHSANASLEVSIVKL 618
 QY 101 EVXGGVII-KVDG-KYVYVKDAHADNVRT-----KDEINRQKO- 138
 Db 619 QVQKVLIIKKIEDLRKIELFLKNAQLKDSIHVFNIXYKQNKPEPYLYLVLRKEVDKLEF 678
 QY 139 -----EHVKDNEKVNWA---VARSGRTTNDG-----YVENPADIIEDT 177
 Db 679 IPVKMKLKEQAVLSITQPLVAASE---TTEDGGHSTHTLSQSGETEVTETETET 735
 QY 178 GNAYIVPHGGHYHYI-----PKSDLSASELAALAKHLAGNKOPOLSYSTASNNQTS 232
 Db 736 -----VGHSTTTVTITPPKESAPKVKVYVENSIEHK-----SNDNSQA 774
 QY 233 VAKGSTSKPANK-----SENLOSLLKELYD-SPSAORYSESDCLVDPAPK 276
 Db 775 LTTVYLLKLUDEFLTSYICHKILVNSMSQOKLLEVYNLPPEERELKS-----CDPLD 830
 QY 277 IISRTPNVAIPHG-----DHFHTFYPSK-----LSALEEK-----IARMVPIEG 316
 Db 831 LLFNQNNIPAMSLYDSMNNDLHLFFELYQKEMIIYLHLKKEENHKKLLEEOQKITG 890
 QY 317 TGSTVST-NAKPNEVSSLSGSSNPSLTTSKELSSADGFIENPKDIVEETATAYIVR 375
 Db 891 TSSTSPGNTVNTAQSATHSNQSOOSNASS---TNTONG-----VAVSSGPVAVEE 940
 QY 376 HGHDFHYIPKSNQI-GOPTLPN-NSLAPSPSLPINPGTSHEKHEEDGYGDFDANRIAE 432
 Db 941 SHDPLTVLSTISNDLKGIVSLNLGNKTKYPNP-LTIST-TEMEKFEYN-----ILKN 990
 QY 433 DESGFVMSGHNNHYFFPKD-----LTBEQIKAAQKHLEEVKTS-----HN-----GL 475
 Db 991 NDTYF-----NDDIKQFVKSKSVITGLTETQKNALNDEIKKLADTLQLSFDLYNKYKLL 1046
 QY 476 DLSLSHEDPYGNAKEMKD---LDKKIEEKIAGIMQYQVKRE-SIWNKRENAII----- 527
 Db 1047 DRFPNKKELGQDKMQIKTLTLEQESLNSLNPHNVLFNSVFFNKKKKEAEIATEE 1106
 QY 528 -----YPHGDHHRADPIDENKPGVI---GHSHSNYELFKPEEGVAKKE 567
 Db 1107 NTLNENTKILLKHYKGLVKNYNGE---SSPLKTLSEVSIQTEDNYANLEKFRALSKIDGKL 1163
 QY 568 GNKYVTGEE-----LTWVNLKNTSTNNONFTLANQKRVSPSFPPELEKGLGINM 619
 Db 1164 NDNLHGLKKKLSLSSGLHLHITELK-EVILKNKYT-----GNSPSNNKK--VNE 1211
 QY 620 LVKL---ITPDGKYLEKSVGKFGVGNFTANFELDPQYLPQGTFRYTI-----ASKDVP 671
 Db 1212 ALKSYENFFPEAKYTVTVTP-----POPDVTPSPLSVRVSGSGSGTKEET 1256

QY 672 EVSYDGTFTVPTSTIAYKWSQTIYPYFPHAGDTYLRVNPQFAVPKGTD-----ALV 721
 Db 1257 QIPTSG-----SLLTQLQVQVOLQNYDEEDDSLVLPFIEGSEDNDEYLDQVVTGEALS 1310
 QY 722 RVFDEFHGNAYLNNYKVGELPIPKLNOGTFTAGNKIP---VTMANA----- 769
 Db 1311 VTMDNILSG--FENEYDVILK-PL-----AGVYRSKKQIEKNIITFNLDNLILASRLK 1363
 QY 770 ----YLD-----NQSTIVEVPILEKENQTDKPSILPQPKRKNK-----A 804
 Db 1364 KRKYFLDVLSDLMQFKHISSNEVYIIE-DSFKLLNSQKNTLLKSYKIKYESVENDIKFA 1422
 QY 805 QENSKLEDKV-----EKPTESEKVEKELSETGNTSTNSLEEVPTVDPQEKVAKF 856
 Db 1423 QEGISYVEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPS-----PAKTDQKRESKF 1477
 QY 857 A-----ESYGMKLENVLFNN-----DGTIELYL 879
 Db 1478 LPFLTNIETLNNLVNKDDYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLF- 1536
 QY 880 PSGEVIKKNMADFTGEAPQNGENKPSSENK-VSTGVENHQP-----ENKPADSLPE 931
 Db 1537 -----KNTNDFEAIKRLINDDTKMDLGLKLLSTGLVQNFPTIISKLEIGKFDQDMLNI 1589
 QY 932 APNE--KPVKPPENS 943
 Db 1590 SHOCVKKQCPENS 1603

RESULT 2
 MSPL_PLAFF STANDARD: PRT: 1701 AA.
 ID P13819;
 AC 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMW5A).
 GN MSP-1
 OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID:5837;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88142999; PubMed=2449612;
 RA Peterson M.G., Coppeel R.L., McIntyre P., Langford C.J., Woodrow G.,
 RA Brown G.V., Anders R.F., Kemp D.J.,
 RT "Variation in the precursor to the major merozoite surface antigens
 of Plasmodium falciparum";
 RL Mol. Biochem. Parasitol. 27:291-302(1988).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 (Potential).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 of the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M19143; AAA29653.1;
 DR PIR: A54498; A54498.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF00008; EGF: 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 FT Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.

```
FT CARBOHYD 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 536 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 607 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193719 MW; 3920875E73D38552 CRC64;

Query Match
Best Local Similarity 4.1%; Score 220; DB 1; Length 1701;
Matches 246; Conservative 171; Mismatches 395; Indels 402; Gaps 65;

QY 3 YALNHRQSEN---KDNRRVSYVDSQSQSKSENLTPOVSKQEGIOAEQIVIRKITDQG 58
Db 519 YTYNVEKQRYNKNFSSNSNVN---QKLKALSYLE-DYSLRKG1-----561

QY 59 YVTSHGCHYHYNGKYPYDA---LFSEEL-----LMKDPNYQLKDADIVN- 100
Db 562 ---SEKDFNHYTLTKGLEADIKKTEETKSSKILEKNFKGLTHSASLEVSIDVKL 618

QY 101 EVKGVII-KVDC--KYVYVLKDAHADNVRT-----KDEINRQKO- 138
Db 619 QVQKVLKIKIEDLRKIELFLKNAQKDSIHVPNIYKPNKPEPYLIIVLKKEVDKLEF 678

QY 139 -----EHWKDNKVNVA---VARSQGYTTNDG-----YVFNPAIDIED 177
Db 679 IPKVKMLKKEQAVLSSITQPLVAASE---ITDGGHSTHTLSQSGETEVTETETET 735

QY 178 GNAYIVPHGCHYHI-----PKSDLSASBLAAKAHLAKGNQPSQLSYSSASTDNNTQS 232
Db 736 -----VGHITVTITLPPKEESAPKVKVYVENSIEHK-----SDNNSQA 774

QY 233 VAKGTSKPANK-----SENLSLKLKYD-SPSAQRYSESGLVDFPAK 276
Db 775 LKTVYVLKLEDFLTKSYCHRYKLYVLSNSMDQKLEVINLTPPEENELKS-----CDPLD 830

QY 277 IISRTNGVAIPHG-----DWHFTPYSK-----LSALEEK-----IARVVPISG 316
Db 831 LFNIONTPAMYSLYDSNIDLOHLFFELYOKEMYIYLHKKEENHKKLEEQKQITG 890

QY 317 TOSTYST-NAKPNEVYSSGLSSNPSSLTTSKELSSASDGYIFNPKOIVETATAYIVR 375
Db 891 TSSTSPGNTVNTAQSAHNSQOQSNASS---TNTQNG-----VAVSSGPAVVEE 940

QY 376 HGDHFIHYKPSNQI-GQPTLPN--NSLATPSPSLPINTGTSHEKHEEDGYGDANRITAE 432
Db 941 SHDPLTVLSINDLKGIVSLNLGNKTKVPNP-LTIST-TEMEKPYEN-----ILKN 990

QY 433 DESGFVMSHGDNHFFKKD-----LTPQKAAQKHLKEEVKTS-----HN-----GL 475
Db 991 NDTYF-----NDIKQFVKSNAVIITGLTQKNALNDEIKKLDLQLSFDLYNKYKLL 1046

QY 476 DLSLSHEQDYPGNKEMKD---LDKKIEPKIAGIMQYGVKRE-SIVNKKENNAII-----527
Db 1047 DRLFNKKKELGQDKWQIKKLTLLKLEQESKLSLNNPHNVLFNSVFFNKKKEAEIATE 1106

QY 528 -----YPHGDHHDADIDHKPVG1-GHSHSNYELFKPEEGVAKKE 567
Db 1107 NTLIENTKILLKHYKGLVKYNYNE---SSPLKTLSEVSTQEDNYANLEKFRALSXIDOKL 1163

QY 568 GNKYVTGEE-----LTVNVLNKNSTFNQNFETLANGQKRVSPFPELEKKLGINM 619
Db 1164 NDNLHLGKKKLSFLSSGLHHLITELK-EVINKKNVT-----GNSPSENKKK--VNE 1211

QY 620 LVKL---ITPDCKVLEKVSQKVFGEVGNIANFELDQVLPQCTKYII-----ASKDYP 671
Db 671
```

```
Db 1212 ALKSVENFLPEAKYTVVTP-----PODVTPLSVRSVSGSSGSKTEET 1256
QY 672 EVSYDGTTVTPTSAYKMASOTIYFPFHAGTYLRVNPQFAVPKGTD-----ALV 721
Db 1257 QIPTSG-----SLLTEQQVVLQNYDEDDSLVLPFIFGESEDNDYLDQVVTGEAIS 1310
QY 722 RVDEFHGNAYLNNYKVEIKLPIKLNQGTTFAGNKIP---VTFMANA-----769
Db 1311 VTMDNILSG--FENEYDVYIK-PL-----AGVYSLKQIEKNIITFNLDINLSRLK 1363
QY 770 ----YLD-----NOSTYIVPEILEKENQDTPKSPILPOFRNKK-----A 804
Db 1364 KRKYFLDVLESDLAQFKHISSNEVIE--DSFKLLNSEQNTLLSKYKVIKESVENDIKFA 1422
QY 805 QENSKLDEKV-----EETSEKVEKEKLSLSTGNSNSTLEEYPTVDPQVEKVAKF 856
Db 1423 QEGISYIEKVLAKYKDDLESIKKVIKEKEKPSPTTPPS-----PAKTDEORRKSKE 1477
QY 857 A-----ESYGMKLENVLFNM-----DGTIELYL 879
Db 1478 LPFLUTNIETLYNVLNVKTDVYLINLAKINDCNVEKDEAHVKITKLSLKAIDOKIDLF- 1536
QY 880 PSGEVIKKNMADFTGEAPQNGENKPSNGK-VSTGTVENQPT-----ENKPADSLPE 931
Db 1537 -----ANTNDFFAIKKLINDDTKKMDLGLLSTGLVQNFPTIISKLEIGKTFQDMLNI 1589
QY 932 APNE--KPVKPPENS 943
Db 1590 SOHQCVKKQCPENS 1603

RESULT 3
CNA_STAAU
ID CNA_STAAU STANDARD; PRT; 1183 AA.
AC Q53654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
Lindberg M., Hoeoek M.;
RT "Molecular characterization and expression of a gene encoding a
Staphylococcus aureus collagen adhesin.";
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
Lindberg M., Hoeoek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoeoek M.;
RT "Identification and biochemical characterization of the ligand
binding domain of the collagen adhesin from Staphylococcus aureus.";
RL Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
Moore D., Jin L., Schneider A., DeLucas L.J., Hoeoek M.,
Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
aureus adhesin.";
```

RL	Nat. Struct. Biol. 4:833-838(1997).	
CC	-!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYOCOCCAL CELLS TO	
CC	COLLAGEN-CONTAINING SUBSTRATA.	
CC	-!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by	
CC	an amide bond (Potential).	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL: M61736; AAA20874.1; .	
DR	PDB: 1AMX; 24-JUN-98.	
DR	InterPro: IPR001899; Gram_pos_anchor.	
DR	TIGRFAWS; TIGR01167; LPXTG_anchor; 1.	
DR	PROSITE; PS00847; GRAM_POS_ANCHORING; FALSE_NEG.	
KW	Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.	
FT	SIGNAL 1 29	
FT	CHAIN 30 1154	
FT	PROPEP 1155 1183	
FT	DOMAIN 151 318	
FT	DOMAIN 533 1093	
FT	DOMAIN 1093 1157	
FT	REPEAT 533 719	
FT	REPEAT 720 906	
FT	REPEAT 907 1093	
FT	SITE 1151 1155	
FT	MOD_RES 1154 1154	
SQ	SEQUENCE 1183 AA; 133066 MW; 66A1CC072E575D76 CRC64;	
	Query Match 3.9%; Score 205.5; DB 1: Length 1183;	
	Best Local Similarity 19.7%; Pred. No. 0.027;	
	Matches 214; Conservative 139; Mismatches 411; Indels 323; Gaps 50;	
QY	16 NNRVSVVDGSSQSKSENLT-PDOVSQKEGIAQAEIVIKITDQGYVTSFGDHYHYNGK- 73	
DB	195 NNEKSYV-----SKDITIKDIOGGQOLDLSTLINVNT-----GTHSNYVSGS 238	
QY	74 --VPYDALFEELLKDPNYQLKADAI-----VNEVKGGYIYIKVDG---KYVYLKDA 122	
DB	239 AITDFEAPGSKITVDNTKNTIDTIPQGYSGNSFSINYKTKIINEOQKEFVNNSQAW 298	
QY	123 HADNRTKDEINQKQEHYKDNKENVNV-----AVARSGRYTIN 163	
DB	299 YQEH--GKEEVNGKSFNHTVHNINANAGIEGTGVKELVKDKDKTAPIANVKFKLSKK 356	
QY	164 DGXYF-----NPADIIEDTGNAYIVPHGGHYHYIPKSDLSASELAAKAHLAGKNMPSQL 219	
DB	357 DGSVVRDQKKEIITDANGI-----ANIKALPSGDYILKETEAPRPYTFDKDKE--- 406	
QY	220 SYSTASDNTQSVAGSKSPANKSENQLSKELLYDPSAQRYSESGLYFDPKAIIS 279	
DB	407 -YPFTKRTDNQGYF--TTTENAKAIEKTDV-----SACK-----WVEGTQYK 448	
QY	280 RTPNGVAIPGHDHYHFTPYSK--LSALEEKIARMVPISGTGSTVSTNAKPNVSVLSGL 337	
DB	449 PTIYFKLYKQDDNQNTTPVDKAIKKLED-----GTTKYVWSNLSPND----- 491	
QY	338 SSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFPYTPKS-NQIGQPTLP- 395	
DB	492 -----KNGKAIKYILVKEVNAQGEDTTPE 514	
QY	396 -----NNSLATPSLPI-NPQTSHEKHEEDGYGFANDRIIAEDSGFYMVSHGDHNYHF 449	
DB	515 GYTKKGLVVTNTEKPIETTSISGEKVDKQDQKGR--PEKSVNLLANG----- 565	
QY	450 KKDLEEQIKAAQKHLBEVKTSHNGDLSLSHEQDYPGNAKEMKDKDKIE-EKIAGIMK 508	
DB	566 -----EKVKT-----LDVTSETNWKY-----EFKDLPKYDEGKKIEYTVT 600	

509 QYGVK-----RESIVVNK-----EKNAILYPHGDHHDHADDIDEHKPGVIGSHSNVLELEK 558

601 EDHVKYDTTDTINGTTITNKYTPGETSATVTKNWDNNNQ--DGKRPTET-----KYVELY- 652

559 PEEGVA-----KKEGNKV-YTGEELTNVYNLLKNSTFNQNET 595

653 -QDGKATGKTALINESNNWTHWTGDEKAKGOOVKYVEELTKV---KGYTHVDND 707

596 LANGQKRVSPFPELEKKILGINNLVKLIITPDGKVLKESGKVFGEG-----VGNIAN 648

708 M--GNLIVTNKYTPETTSISGEKVDKQDQKRPKESVNNLLADGKVKTLIDVTSETN 765

649 FELDQPLP-----QOTKTYIAS--KDYPEVSVGDTFTVPTSLAYKMASOTIFYPFHAG 701

766 WKYEFKDLPKYDEGKKIEYTVTHDVKADY--TTDINGT-----TITNKYTPG 810

702 DTYLRV-----NPQFAVPKGTDALVRVDE-----FHGNAYLENNKYV 739

811 ETSATVTKNWDNNNQDGKRPTETIKVELYQDGKATGKTALINESNNWTHWTGLDEKAKG 870

740 GEIKLPIPKLNQ---GTRTAGNKIPVTFMANAYLDNQSYIVVEPILEKENOTDK--- 792

871 QQVAYTVEELTKVGYTHVDNDNMGNLIVTNKYTPETTSISGEKVDKQDQKRPK 930

793 -PSILPQFKRKA-----QENSKLDE--KVEEPTKSEKVEKLESETGNST 835

931 VSVNLLANGKVKTLDTVSETNWKYEFKDLPKYDEGKKIEYTVTHDVKDYTTDINGTTI 990

836 SNSTL--EIVPTVD-----PVEKVAKFAESYGMKLENVLFNMDGTIELYLP 880

991 TNKYTPGETSATVTKNWDNNNQDGKRPTETIKVELYQDGKATGKTALINESNNWTHWTG 1050

881 SGEVKKNMADFT-GEAPOGNENKPSKSVGTCTVENQPTENKPADSL-PEAPNEK-- 936

1051 LDEKAKGOOVKYTVDELTKVNGYTHVDNDNMGNLIVTNKYTPKPKPKPIYEPKPKDTP 1110

937 PVKPENS 943

1111 PTKPDHS 1117

RESULT 4

IGA4_HAEIN STANDARD: PRT; 1849 AA.

AC P45386;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).

GN IGA.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NHU1 HK61;

RA MEDLINE=92234949; PubMed=1373717;

RA Poulson K., Reinholdt J., Kilian M.;

RT "A comparative genetic study of serologically distinct Haemophilus

RT influenzae type 1 immunoglobulin A1 proteases.";

RL J. Bacteriol. 174:2913-2921(1992).

CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A

CC -!- PRODUCING INTACT FC AND FAB FRAGMENTS.

CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at

CC certain Pro-I-xaa bonds in the hinge region. No small molecule

CC substrates are known.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC

CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE

CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE

CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY

CC SIMILARITY).

```

CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC
DR  EMBL: M67491; AAA24968.1; -.
DR  MEMOPS: S06.001; -.
DR  InterPro: IPR000710; IGA_S6.
DR  InterPro: IPR004899; Pertact_sup.
DR  Pfam: PF02395; IGA1; 1
DR  Pfam: PF03212; Pertactin; 2.
DR  PRINTS: PRO0921; IGASERPTASE.
KW  Hydroxylase; Serine protease; Transmembrane; Zymogen; Signal.
FT  SIGNAL 1 25
FT  CHAIN 26 1021
FT  PROPEP 1022 1849
FT  ACT_SITE 299
FT  SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;
SQ
Query Match 3.8%; Score 203.5; DB 1; Length 1849;
Best Local Similarity 19.3%; Pred. No. 0.062;
Matches 239; Conservative 153; Mismatches 449; Indels 397; Gaps 59;
QY 8 HRSQENKNNRYSYVDGSSQSKSENLTLDQVSKQEGIQ-----AEQIVIKINDQ 57
DB 125 HRDVSSENR---YTVENKNFTFNTVFTTKEDQAQRREDYMPPLDFVTEVAPI 181
QY 58 GYVTSIGDHYHYNGKVPYDALF-----SEELMKDPNQOL-----KDAD---IVNEVKG 105
DB 182 EASTANNKGE-YNNSDKYPAFVRLGSGSQFIYKKGSRVQLILTEKDKGNLLRNWDVGG 240
QY 106 YIKVDGKYVYVLKAAHADNVRTKDEINRQKHVKNKNSVNAVARSQGRITINDG 165
DB 241 DNLELVGNAYTY-----GLAGTPYKVNHEN-----NGLIGFNSKEEHSDDPKG 283
QY 166 YV-----FNPADIETDGNAYIV--PHGGHYHYIPKSDL-----SASLAATAKHLAKG 212
DB 284 ILSQDPLTNVAVLGDGSGFLFVVDREKGNWFLGSDFWAGYKKSQWENNIYKHEPAEK 343
QY 213 NNQPSOLSYSS-TASDNNTQSV--AKGSTKSPANKSENLSQSLKELYDPSAQRY----- 264
DB 344 IYO-----QYSAGSLTGSNTQYTWQATGTSITITGCEPLSVDLTDGKOKPNRHKGSITLKG 399
QY 265 -----SESQGLVFDPAKIIISRTPN-----GVAIPHGD-----HYHFIPYSKLS 302
DB 400 SGTLLNHNIDOGAGGLFPEGDYEVKGTSDSTTKGAGVSVDGKTVTKVKNPKYDRLA 459
QY 303 -----ALEEK-----IARMVPISGTGSTVSTNAK- 326
DB 460 KICKGTLVVEGKNGEGLLKVGDTGVTILKQKADANKKVAQFSGVIGSRSTLVINDKQ 519
QY 327 --PNEVSSSL--GSLSSNPSSL-----TTSKELSSADGYIFNEKD 363
DB 520 VDPNSIYFGRGRDLNGLNSLTFDHIIRNIDGCAVNVNHNMTNINITGSLTNP-- 577
QY 364 IVEETATAVIVRHGDHF-----YIPKSNQIQGQTLPLNNSLATPSPSLINPG 411
DB 578 ---NTITSYNIEAQDDHPLRIRSPYRQLYFNQDN--SYVTLKKG--ASTSESLPQNSG 631
QY 412 TSHEKHEEDGYGFD-----ANRIATAEDSGFVMSGHGDNHVFYFKDTEBQIKAAQKHL 465
DB 632 ESENWNLWYGRTSDBAKRNVNMHNERNNGF-----NGYF-----GEETKATQN-- 677
QY 466 ESVKTSNGLDLSLSHEQDY--PGNAKEMKDLDKKIEEKIAGIMQYGVKRESIVVNKEK 523
DB 678 GKLNVTENG-----KSDQNFLLTGTGNTLNGDL-----NVKEK 709
QY 524 NAIITPHGDHHDHPIDEHKPVGIG-----HSHSNYELFKPEEGVAKK-----EGN 569

```

```

DB 710 GTLFLSGRPTPHARDI-----AGISSTKKDHFHTENNEVVVDWNNFRKATTMNVGN 764
QY 570 -KVITGEELTNVNVLLKNSTNNQNFLLANGQKRVSFPPPELBKKGILNNLVK----- 622
DB 765 ASLYSGR--NVANITSNITASN-----NAQVILGY-----KTGDTVCVRSDDITGY 807
QY 623 LITPDCKVLEKSVKVFEGCV-GNI-----ANFELDQYLPQGFKYTIASKDYPEVSYD 676
DB 808 VTCNSNLSSEKALSNFNTNLRGNVNLTEASFTLGRKANLFG-----TIQSIGTSQVN-- 860
QY 677 GTFTVPTSLAYKMASQTIFFPHAGDTVLRVNPQFAVFKGTDALVRVDFEHGNA----- 731
DB 861 ---LKENSHWHLTGNVNVNQLNLIINGHILHNAQDANKVTTYNTLVNSLGSNGCSFYW 916
QY 732 -----YLENNYKVEIKLPI-----PKLNQGT-----TRTAGNKIPVTMANAYL 771
DB 917 VDFTNKSNKVNVKASATGNFTLOVADKGTGEPNHNELTLDASNATRNLEVT-LANGSV 975
QY 772 DN-----QSTYIIVEVPILEKENOT-----DKPSILPQFKRNKAQEN 807
DB 976 DRGAKYKLRNVNGRYDLYNPEVEKRNQTVDTTNTITPNDIQADAPS-----ROSN 1026
QY 808 SKLDEKVEEP-----KTSEKVEKEKLSETGNST 835
DB 1027 NEELIARVETPVPPAPAPATESAIASEQPETRPAETAQAQPAEMEETNANSTETAPKSDTATQT 1086
QY 836 SNSLTLEEVPTVDYQKQVAKFAESYGMKLENVLFNMDGTIELYLPSEVYIKKNMADF--- 892
DB 1087 ENPNSSEVPS-----ETTEKVAENPPQENETVAKNEQATEPTQNGEVAKEADQPTVEAN 1141
QY 893 --TGAPGCGNENKPSNGKYSTGTVENQPT-----ENKPADSLPAPAPNEKPV-----K 939
DB 1142 TOTNEATQSECKTEETQ-----TAETKSEPTSVTSVENQPEKTVSQSTEDKVVVKEEK 1196
QY 940 PENSTDNQMLNPEGNVGSDDP-MLDPALPEEAPAVDPVQE 976
DB 1197 AKVETEETQKAPQVTSKEPPKQAPPAPEVPTDTNAEE 1234
RESULT 5
MSPL_PLAFC STANDARD; PRT: 1726 AA.
AC P04934;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5835;
RN [1]
RX MEDLINE=96205236; PubMed=3517809;
RA Weber J.L., Leisinger W.M., Lyon J.A.;
RT "Variation in the gene encoding a major merozoite surface antigen of the human malaria parasite Plasmodium falciparum."
RL Nucleic Acids Res. 16:1206-1206(1988).
CC [2]
CC SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).
CC -!- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa and 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----

```


CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M37213; AA29611.1; --
 CC InterPro: IPR000361; EGF-like.
 CC Pfam: PF00008; EGF; 1.
 CC Malaria: Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19
 FT CHAIN 20 1726
 FT CARBOHYD 133
 FT CARBOHYD 272
 FT CARBOHYD 501
 FT CARBOHYD 567
 FT CARBOHYD 638
 FT CARBOHYD 827
 FT CARBOHYD 924
 FT CARBOHYD 944
 FT CARBOHYD 990
 FT CARBOHYD 1016
 FT CARBOHYD 1114
 FT CARBOHYD 1221
 FT CARBOHYD 1613
 FT CARBOHYD 1726
 FT SEQUENCE 1726 AA; 196174 MW; 5B59CEBFA2F9A026 CRC64;
 Query Match 3.7%; Score 199; DB 1; Length 1726;
 Best Local Similarity 19.7%; Pred. No. 0.091;
 Matches 239; Conservative 168; Mismatches 402; Indels 402; Gaps 63;
 QY 3 YALNHRSEN---KDNKRVYVDGSSQSSOKSNTLPDQVSKQEGIOAQEIVIKITDQG 58
 Db 550 YIYNVEKQYNNKFFSSNNVNV---OKLKAALSVLED-YSLRGI----- 592
 QY 59 YVSHGDHYHYNGKVPYDA---LSEEL-----LMKDPNYQLKDAIYN- 100
 Db 593 ---SEKDFNYHTLTGLEADIKLTETEEKSSSENKILEKNFKGLTHSANASLEVYDVKL 649
 QY 101 EVKGGVII-KVDG--KYVYVLKDAAHADAVRT-----XDEINRQKQ- 138
 Db 650 QVQKVLIIKIEIDRLKIELFLANQKLSIHVPNIYKQNKPEPYIYLVKREVDKLEF 709
 QY 139 ----EHVXDNKVNVA---VARSQRYTTNDGIVFNPADIIEDTGNAYIVPHGGHYH 190
 Db 710 IPKVKMLKKEQAVLSITQPLVAASE---TTEDG-----GHSTH 746
 QY 191 YIPKS-----DLASFLAAKAHLAKGNMOPQSISYSTA---SDNNTQSVAKGSTS 239
 Db 747 TUSQGETVETETEETETVGHHTTITLPPKVKVYVENSIEHKSNDNSQALKTIVYL 806
 QY 240 KPANK-----SENLOSLLKLYD-SPSAGRYSESDGLVDFDPAKIISTRPN 283
 Db 807 KILDELFTKSYICHKYLIVNSNMQKLLVYNLTPPEENELKS---CDPLDLIFNQIN 862
 QY 284 GVAIPHG-----DHYHFTPYK-----LSALEEK-----TARMVPTSGTGSTVST 323
 Db 863 NTPAMYSLDYSNMNDLQHLFFELYQKEMIIYLLHKLKEENHKKLLEEQKQITGTSTSSP 922
 QY 324 -NAKPNVYSSLSGSSNPSSLTTSKELSSASDGYIFNPKDIVEATATAYIVRHGDHRYH 382
 Db 923 GNTVNTAQSATHSNQNSQNSASS---TNTONG-----VAVSSGFAVVEESHDPLTV 972
 QY 383 IPKSNQI-QQPTLPN--NSLATPSPSLINPGTSHKHEEDGYGFDANRIAEDESGFVM 439
 Db 973 LSTSDNLKGIIVSLLNGNKTVPNP-LTIST-TEMEKFFVN-----ILKNNDTYF-- 1020
 QY 440 SIGDHNHVFPKD-----LTEQTKAAQKHLEEVKTS-----HN-----GLDSLSSHE 482

Db 1021 --NDIKOFVSNKSVITGLTETQKNALNDIKKLTQLSFLLNKYKLDRLFNKK 1078
 QY 483 ODYPGNAXEMKD---LDKKIEIKIAGIMKQYGVKRE-SIVVNKEKNAIL----- 527
 Db 1079 KELGQDKMQIAKLLTKLBQLESKLSLNPHNVQLNFSVFNKKKEAEIAETENTLENTK 1138
 QY 528 -----YPGDHHHAADIDHEKPVGI--GHSHSNYELFKPEEGVAKKEGNKYVTG 574
 Db 1139 ILLKNYKGLVYYNGE---SSPLTKLSEVSIQTEDNANLEKFRVLKIDGKLDNDNLHLG 1195
 QY 575 BE-----LTNVNLLKNTFNQNFNFTLANGQKRVSFSPFPELEKKLGINMLVK 622
 Db 1196 KKKLSFLSSGLHQLITELKEVINKNKNYTGNS-----PSENKKK--VNEALK 1239
 QY 623 L---ITPQGVLEKVSQKVGFGVGNNTANFELDOPYLPQGTFKYTI-----ASKDYPEVS 674
 Db 1240 SYENFLPEAKYTTVVTP-----POPDVTIPSLSVRSVSGSSGSTKEETQIP 1284
 QY 675 YDGTFTVPTSLAYKMASOTIFYPPHAGDTYLRVNPQFAVPKGTD-----ALVRVF 724
 Db 1285 TSG-----SLTLELQVQVQLQNVDEEDDSLVLPITGSESDNDEYLDQVVTGEAIVTM 1338
 QY 725 DEFHGNAYLENNYKVEIKLIPKLNQQTATTAGNKIP---VTMANA----- 769
 Db 1339 DNILSG--FENEYDVIYK-PL---AGVYRSLKKQIEKNIFTFNLDNLINDILNSRLKRRK 1391
 QY 770 YLD-----NOSTYIVVEVPILEKQNDKPSILPQFKRNK-----AQEN 807
 Db 1392 YFLDLESDLQWFKHISSEYIIE-DSFKLLNSQKNTLLKSYKIYKESVENDIKFAQEG 1450
 QY 808 SKLDEKV-----EPEKTSKEVKEKLSGTGNTSTNLEEYVTPDVPQEKVAKFA-- 857
 Db 1451 ISYVEKVLAKYKDDLESTKKVYKEKEKFPSPPTTPS-----PAKTDEQKESKELPF 1505
 QY 858 -----ESYGMKLENVLFNM-----DGTIELYLPSPG 882
 Db 1506 LTNETLYNNLVNKDDYLLNLKAKINDCNVKEDEAHVKITPKLSLDLRAIDDKIDLF-- 1561
 QY 883 EVIKKNMADFTGEAPQNGENKPSENGK-VSTGTVENQPT-----ENKPADSLPEAPN 934
 Db 1562 ---KNHNDFAIRKLINDDTKMDLGLKLLSTGLVQNFNPNTIISKLLIEGKFDQMLNISQH 1617
 QY 935 E--KPVKPENS 943
 Db 1618 QCVKRCQCPENS 1628
 RESULT 7
 MRSP_STAAU STANDARD; PRT; 1637 AA.
 ID MRSP_STAAU STANDARD; PRT; 1637 AA.
 AC P80544; Q92P62;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Methicillin-resistant surface protein precursor.
 GN PLS.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Isolate 1061.
 RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.,
 RT "Pis, a large repeat-rich surface protein of methicillin resistant
 Staphylococcus aureus";
 RL Submitted (DSC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 624-628; 676-682; 938-948; 1156-1168; 1176-1185;
 RP 1199-1205 AND 1217-1224.
 RC STRAIN-Isolate 1061.
 RA MEDLINE=96270743; Pubmed=8665912;
 RX Hilden P., Savolainen K., Tyyneä J., Vuontola M., Kuusela P.;

"Purification and characterization of a plasmin-sensitive surface protein of *Staphylococcus aureus*," Eur. J. Biochem. 236:904-910(1996).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (Potential).
 CC
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF115379; AAD09131.1;
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor; 1
 DR TIGRFAMs: TIGR01167; LPXTG_anchor; 1;
 DR TIGRFAMs: TIGR01168; ISIRK_signal; 1;
 DR PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
 KW Antibiotic resistance; Glycoprotein; Cell wall; Peptidoglycan-anchor; Repeat; Signal.
 FT SIGNAL 1..48
 FT CHAIN 49..1601 METHICILLIN-RESISTANT SURFACE PROTEIN.
 FT PROPEP 1602..1637 REMOVED BY SORTASE (POTENTIAL).
 FT DOMAIN 1301..1582 141 X 2 AA TANDDEM REPEATS OF D-[SAG].
 FT SITE 1598..1602 LPXTG SORTING SIGNAL (POTENTIAL).
 FT MOD_RES 1601..1601 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 SQ SEQUENCE 1637 AA; 174573 MW; 75BESAD84693D309 CRC64;
 Query Match 3.7%; Score 195.5; DB 1; Length 1637;
 Best Local Similarity 19.0%; Pred. No. 0.12;
 Matches 222; Conservative 138; Mismatches 454; Indels 357; Gaps 48;
 QY 2 AYALNQRSEKDNRRVSVGSG-----QSSCKSENTPQVSOKEIQAEQIVKIT 55
 DB 46 AEAANNITQKDDSDASAKVKNVGTIEQSSANSNESDIPEQVD-----VTQDT 95
 QY 56 DQGVYTSKGDHYHYNGKVPYDALFSEELMKDPYQLKADAVNEVKGYYIKVDGKYV 115
 DB 96 TE-----QASTEKANTTEQASTEETKADTEQATTEAPKAEQTDK 136
 QY 116 VYLKDAHADNRVTKDEINQKQEHVNDKNEVNSVAVARSQGRYTNDGYVFNADIEE 175
 DB 137 VETEEAPKAEET---DKATTEAPKAEETDKATEEAPKTEETDKATTEE-----APAAE 187
 QY 176 DTGNAYIVPHGHVHYLPKSDLS-----QASTEKANTTEQASTEETKADTEQATTEAPKAEQTDK 198
 DB 188 ETSKAAT-----EAPKAEETSKAATEAPKAEETKATEEAPKTEETDKVETEAP 240
 QY 199 -ASELAALAKHLAKNQPSQLSYSTASDNTQSVAKGSTSKPANKSENLOSLLK---- 253
 DB 241 KAEETSKAATEKAPAEETNKVETEAPAEETKAAETEET--PAVEDTNAKSNSNAOPS 298
 QY 254 -----ELYDPSAQRYSESDGLVDEDAKISRTPNGVALPHGDHYHFIPYSKLSALE-EK 307
 DB 299 ETERTQVVDIVAKDLKSKSEVTEAKAEKVLPKDI-----SNLSNEETKK 345
 QY 308 IARVPISTGTSVSNKAPNEVYSSLSNPPSLTSSKELSSAGSYINPKDIVEE 367
 DB 346 IALSEVLKETAN--KENAQPRATFRSVSSNARTTNVYSATALRAAA-----QDTYTK 396
 QY 368 TATAYIVRHGDHFIYPKNSQIGOPTL-----PNNSLATP-----SPSL 406
 DB 397 KGTCNFTAHGDIHKYKEEFPNEGTLTAFNTFNPTGKCALEYNDKIDNKFDTITV 456
 QY 407 PINPTGSHKHEEDGVGFANDRIAD---ESGVNSHGCHNHFFKDLTTEQIKAAQK 463
 DB 457 PVANNQNTGTGADGWFMTQCGNGDFLNOGGILRDKGMANSGFKIDTAYNNYNGVD 516
 QY 464 HLEVTKSHGLDLSLSSEHODYPCNKAENKDKLKEERAGIKQYGVKRESIVVNEK 523
 DB 517 KLDADKI--NNLSQIGAAKGVGYCTFKVNGAD-----GVTVQVG---QNALNTRDK 561

QY 524 --NALLYPHGDHHRADPIDEHKPVGIGHSHS-----NYELEKPEEGVAKKEGKVIT 573
 DB 562 PVNKITITADNTNHLN-----GFHQGLNDVNLNYD-----AATSTITATYA 604
 QY 574 GEELTNVNLKLNSTFNNQNTLANQKRVFS---FPPELEKKLGNNMLVLIIPDQKV 630
 DB 605 GKWKATTDGLDKSKQYNFLITSSHMQRYSNGIMRTNLEG-----VTITPQADL 657
 QY 631 LEKVGSGVFGESVGNANFELDQYLPQGTQKTYTASKDYPEV---SYDG--TETVPTS 684
 DB 658 IDDV--SVTKQPIPHKTIREFDFTLEPGS-----PDVIVQKGDGEGKTTTTPTT 703
 QY 685 LAYKMASQTIYFPHAGDYLVRNPQFAVPKGG--TDALVRVF-----DEFHGNAYLE 734
 DB 704 --KVDPDT-----GDVVVERGEPTTEVTKNPVDEIVHFTPEEVPOGHKDEFEN--- 749
 QY 735 NNKYVGELKPIPKLNOCTRTAGNKIPVTFMANAVLDNQSTYIIVEPILK----- 786
 DB 750 -----LPI-----DGTEEVPG-----KPKIKNPETGEVVPVDDVTKHGPKA 787
 QY 787 -ENOTDKPSILPQKRNKAQENSKLDEKVEPKTSEKVEKELSETGNSNSTLEEY-- 843
 DB 788 GEPEVTKEEIPEKKEKREFNPKLPGKEKV---TQEGQIGKTTTTPTTINPLTGEKVG 843
 QY 844 --PTVDPVQKAKFAESYGMKLE-----NVLFNKDGTIELYLP-----SGEVI 885
 DB 844 GEPTVTKRPVDEITQFGGEEVPOGHKDEFDNLDPIDGTEE--VPGKPGIKNPETGEV 901
 QY 886 KKNMADFTGEAPQGN-----GENKPSSENGKYSTGV----- 916
 DB 902 TTPVDDVTKHGKAGEPEVTKKEIPEKKEKREFNPKLPGKEKVTEGQIGKTTTTPTT 961
 QY 917 -----ENQPT---ENKPADSLPEAPNEK-----PVKPEKSTDN-----GML 949
 DB 962 NPLTGEKVGEGEPTTEVTKRPVDEITQFGGEEVPOGHKDEFDNLDPIDGTEE-- 1021
 QY 950 NPE--GNVGSDFMLDPALEAPAVDPVQEKLE 979
 DB 1022 NPETGEVVPVDDVTKHGKAGEPEVTKREE 1052
 RESULT 8
 YQ38_CAEEL STANDARD; PRT: 1531 AA.
 ID YQ38_CAEEL Q09459; Q09513;
 AC Q09459; Q09513;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C09G5.8 in chromosome II.
 GN C09G5.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Palmer S., Sulston J.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RA Durbin R.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO C. ELSEGANS ZK666.1 AND ZK666.4
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC


```
CC -----
DR EMBL; 246791; CA86762.2;
DR EMBL; 246792; CA86762.2; JOINED.
DR EMBL; 246792; CA86767.2;
DR EMBL; 246791; CA86767.2; JOINED.
DR WormPep; C09G5.8; C825774.
KW Hypothetical protein.
FT DOMAIN 384 394 POLY-SER.
FT DOMAIN 395 398 POLY-ASN.
FT DOMAIN 1150 1154 POLY-ASP.
SQ SEQUENCE 1531 AA; 174862 MW; A453990AF0DEBC2F CRC64;

Query Match 3.6%; Score 189.5; DB 1; Length 1531;
Best Local Similarity 19.2%; Pred. No. 0.21;
Matches 197; Conservative 160; Mismatches 441; Indels 229; Gaps 39;

QY 113 KYYVYLKDAHADNVTKDEINRQKHVKON-----EKNVSNVAVARSQGRYTNDGYV 167
DB 170 KYNLRDRTSSSDRSARYPIEKWRPOLEDHFHNWVELNAKQAKKVEQEKQIITNSR 229
QY 168 FNPADIEDTGNAVIVPHGGHYHYPKSDLSASALAHAHLAGNMOPSQLSVSTASD 227
DB 230 FRSMLEKRSQNEKVVERSKYDDVKEQNLDMKLAQAQQL-----LIYTAPSA 280
QY 228 NNTQSVAKGSTS-----KPAKNGENLOSLLKELYDSPAQYS--ESDGLVFDPA 275
DB 281 ATTASMTGRSTFRPPSTFRQRPILTATGTSI-----DRPGSAPVARKKSDG---GE 331
QY 276 KIISTPNVAVPHGDHYHIFYSKLSALEKIAWVPSGTGTVSTVNAKPNVYVSLG 335
DB 332 KQLATDEKLAIVRLNRLKNKNDIEITLKYIEKLRLK-----SSVNQSSPTRLSTSS 388
QY 336 SLSSNPSSLTTS-----KELSSADSGYIENPKDIVEETATAYVRHGDHFHYIPKSNQ 388
DB 389 SKSSSSNNNDGEGKDSLEENSESDSGRSTPVIEKKKPR-RKSKKSSHQPSKNP 447
QY 389 IQOPTLPNNSLATPSPPIPGTSHKHEEDCYCFDANRIIAEDSEGVSHGDHNYIF 448
DB 448 IPPPRIPDQTEKVLKDKLKVASNDLAMIQECDLVKKANERLVHOSLSKSTEYGARESIE 507
QY 449 FRKDLT--EEQIKAAQKHLVEEY---KTSHNGDLSLSHEQDYPGNAYEMKDLKKIEEK 502
DB 508 EKKKIVLEELKTEKRIKESHRERREDQKKFEARLH-----YKNKYDAAKTEKK 560
QY 503 IAGIMKQYGVKRESIVNKEKNAIYIPHGDDHHADPIDEHKPVGIGHSHNVLEPKPEG 562
DB 561 LSVAKNSKVEEERI--EEKIS-----HSPPTMTFPIRRKHSQSEISRRMRADD 609
QY 563 VAKKECNKYVTGEELNV-----VNLLKNSTENNQNFTLANGQKRVSPFPELEKKLGI 617
DB 610 DILQ---KLY--KEVADILHSDVGIAEINTLGASENSLARKWKLYSELY-EELEKVR-- 661
QY 618 NML-----VKLITPDGVKLVKGVKGVGEGVGNIANFELDQYLPQGTIFYT 664
DB 662 NMLLIQYDINQOMKEIKLLKDLRLKTVSAEILSK-----SREEVEE-----PQKKIF 711
QY 665 IASKQYPEVSYDG-----IFTVPTSLAYKM-----ASQIFYPFHAGD 702
DB 712 MLEQIRTIAYSGQPKVLLANQIPIPRVNTDLSVKLINVKPSPLSTSFPPFSLEFFD 771
QY 703 TYLRVNPQFAVFKGTALVRVDEPHGNA---YLENNYKVGVEIKLPIPKLNGQITRTAGN 759
DB 772 FQLEITTPMDAKQHNMDFTTVYDVLVSNLLIHYLQTNQIVIEVMPASDCVKLLAAATIS 831
QY 760 KIPV-----TFMANAYLDNGSTVIVEVPILEKENQTKDKPSILDPQKRNKAQENSK--- 809
DB 832 LIPLFEDSVLRFCSIEMLKSDYTG-VEMCTLRVIEVSQP-ISDSFKFKKSEMARNNL 889
QY 810-----LLEKVEEPKTSKVEKEKLSNGSTNSLTLEE- 842
DB 890 PLOLENEDETDNFDPLTMWNVVGLTFGKDPSTECIVDEFISFPYPTDFTSSEI 949
QY 843-----VPTDPVQEKVAKPAESYGMKLENVLFNMGTI-ELYLPSEGVIKN----- 888
```

```
DB 950 RSKRDCYIPKIDIIARNLFATSSISFFEL-IENIP-RQDGVIAITLHLPLCKLGSGIKGT 1007
QY 889 --MADFTG-----EAPQNGENKPKSENGK-----VS 912
DB 1008 FPLMDTGRSPSSVSLDLCLINKHEIPSFLLKHEPKELKEVKDTPILPQPVRRTSKFEVV 1067
QY 913 TGTVENQPTENKPADSLPEAPNEKVPK-----ENSTDNGMLNPEGNVGSDPMLDPALE- 966
DB 1068 TPVKEAEILHDAEPTSMPPKAP-EPTTAPLRRLSTSDSDTSFSSKSDLFSPTNPQTYDY 1126
QY 967 EAPAVDP 973
DB 1127 EIPAVTP 1133

RESULT 9
P3K1_DICDI STANDARD; PRT; 1570 AA.
AC P54673;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Phosphatidylinositol 3-kinase 1 (EC 2.7.1.137) (PI3-kinase)
DE (PtdIns-3-kinase) (PI3K).
CN PI3A OR PIK1.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NBLTAXID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96009592; PubMed=7565716;
RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.;
RT discoidium: biological roles of putative mammalian p110 and yeast Vps34p PI 3-kinase homologs during growth and development.;
RL Mol. Cell. Biol. 15:5645-5656(1995).
CC -! CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol = ADP +
CC 1-phosphatidyl-1D-myo-inositol 3-phosphate.
CC -! SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U23476; AAA85721.1;
CC DictyDb; DD01099; pika.
CC InterPro; IPR002420; PI3K_C2.
CC InterPro; IPR000341; PI3K_ras_bind.
CC InterPro; IPR001263; PI3Ka.
CC InterPro; IPR000403; PI3_P14_kinase.
CC InterPro; IPR000626; Ubiquitin.
CC Pfam; PF00240; ubiquitin; 1.
CC Pfam; PF00454; PI3_P14_kinase; 1.
CC Pfam; PF00613; PI3Ka; 1.
CC Pfam; PF00792; PI3K_C2; 1.
CC Pfam; PF00793; PI3K_C2; 1.
CC Pfam; PF00794; PI3K_rbd; 1.
CC SMART; SM00142; PI3K_C2; 1.
CC SMART; SM00144; PI3K_rbd; 1.
CC SMART; SM00145; PI3Ka; 1.
CC SMART; SM00146; PI3K; 1.
CC PROSITE; PS00915; PI3_4_KINASE_1; 1.
CC PROSITE; PS00916; PI3_4_KINASE_2; 1.
CC PROSITE; PS0290; PI3_4_KINASE_3; 1.
CC Transferrase; Kinase; Multigene family.
KW DOMAIN 41 48 POLY-ASN.
FT DOMAIN 59 73 POLY-ASN.
FT DOMAIN 161 168 POLY-ASN.
```

FT	DOMAIN	294	304	POLY-THR.	
FT	DOMAIN	308	324	POLY-THR.	
FT	DOMAIN	413	424	POLY-ASN.	
FT	DOMAIN	503	510	POLY-SER.	
FT	DOMAIN	570	579	POLY-ASN.	
FT	DOMAIN	821	828	POLY-ASN.	
FT	DOMAIN	831	836	POLY-ASN.	
FT	DOMAIN	1309	1570	P13K/P14K.	
SQ	SEQUENCE	1570	AA; 178374	MW; 558678B72B34D783	CRC64;
Query Match					3.5%; Score 186.5; DB 1; Length 1570;
Best Local Similarity					17.6%; Pred. No. 0.3; 361; Indels 461; Gaps 48;
Matches 205; Conservative 141; Mismatches 141; Indels 461; Gaps 48;					
QY	128	RTKDEINQKOE	---HVKDNEKV	---NSNVAVARSOGRYTNDGYVFNPAADIIEDTIGNA	180
Db	77	KKKDSKQNGEINQEMSENKIIYNSDNCSSGSSGGHVNNGHHI	-----LIEEN	---	128
QY	181	YIVPHGHHYIPKSDLSASELAALAKLAGKNQPSOLSSTASDNTOSVAKGSTS	---	239	
Db	129	-----ERLEHENQIEQIYKOK	---GMEFQKKDLREGYDVNSNNNNNGGSSG	175	
QY	240	-----KPAKSEN	-----LQSLKEL	-----YDSPSAQRYSES	267
Db	176	SSSGSDSASNOPIRTNRREGSILNLKKOGLYKEISQRFQTPDTASYTRPNANNISIK	235		
QY	268	DGLVF	-----DPAKIISRTPNGVAIPGHGDHYHIFPYKLSALEEKIAR	310	
Db	236	DKISILKREQERRKQDSEVOQREKIVIVLSADSNIIQIYHPVSLIEKMKSKLDTEEKPAIT	295		
QY	311	MVPISTGTVSTNAKPEVYSSIGLSNPSL	---TTSKELSSASGDGYVFNPKDIVE	366	
Db	296	TTTTTTTSTSTPTTTTITTTTSTNDITIKPKTSPTKNNEERSQSPITPKQPV	355		
QY	367	ETATAYIVRHGDHFIYPKSNOIGOPTLPNLSLATPSPSLPINQTSHEKHEEDCYGDA	426		
Db	356	E	-----IVKKVS	-----TPKSN	398
QY	427	NRIIADESGFVMSGDHMH	-----	-----YF	448
Db	399	GNLVIYVNTNTSTNNNNNNNNNNNNETIKRGRVLYTPSPDLKKNQIYFTIPINPV	458		
QY	449	-----FKKDLTEEQI	-----	-----KAAQK	463
Db	459	NKTNPQLLSNTSQOFLKTLISNIPDCKINDINDTDAFSDLSASASSSFIRKSSQS	518		
QY	464	-----HLEEVKTSNGLDLSLSSHEQDYPGNAKEM	-----KDLKKIEEKIAGIMK	---	508
Db	519	LLNVQSLRVKAKTSPNIFLM	-----PNOSKILQVKGSDTIENLKERIISDYLFNN	571	
QY	509	-----QYGV	-----KRESIVVVK	-----ERNALIY	528
Db	572	NSNNNNNCKYGADSYLILDFNDNPMERSVLNLSKSDYILDKRAQGLPKLKVIERKSTIL	630		
QY	529	PHGDHHDAPIDHKVPQVIGCHSHNSVELEK	-----PBEQVAKKE	-----	567
Db	631	-----DSDPDSLEP	-----SEVEIIRKLIPGDTWRGEVEYFRVTRGLRYEALP	677	
QY	568	-----GNKYVTG	-----BELTNVWLLKNST	-----FNNQNFILIA	597
Db	678	LKSGTQSTILLVPLSPPIVGNKLLISLIFLPITQVTKTLDLELNATADQFTNRLFT	735		
QY	598	NGQKRVSFPPPELEKKGILGNMLVLIITPDGKVLKRVKGVFGEGVGNATANE	LDQPYL	656	
Db	736	---KNYSKHLN	-----VNSNHFTLKVVGSSDFIHGPHDITRTFESIRNHII	778	
QY	657	PGQTFXYITASKQYPEVSYDGTFTVTSLAYKMASQTIYPFHAGDTYLRVNPQAVPKG	716		
Db	779	QGTQPLTILQRPKPELDPO	-----	-----PF	809
QY	717	TDALVRVDFEFGHNAYLENN	-----	-----YKVGKILPIPKLNOGTR	755
Db	810	L	-----IIDSCHSCNAINCNNNNNTSTNNNNINFDNDWDQITHISIREIKKPFVRKVMGSTR	864	

QY	756	TAGNKIPVTFMANAYLDNQSTVIVEVPILK	-----ENQTDKPSILPQFK	---	800
Db	865	-----IPLSCIKDI	-----DSSSVIVSISLYHGIECFSKAFTQIIPPPFAFLAETLSVDNC	916	
QY	801	-----RNKAQENSKLDERVEEPTSEKVEKLSERFSGNSTSTLEEVPTDVPQVEKVAK	855		
Db	917	EWLVFTNIDYSNLPVDARL	-----SISVYSANETVDDVEE	---	951
QY	856	FAESYGMKLENVLFNMDGTIELYLPSPGEVIKKNMADFTGEAPQNGENK	---PSENGKYST	913	
Db	952	-----IKNLDEATKKLTPIG	---WINVMTITDKYQLRQGMVLSLWPSDFS	997	
QY	914	GTVENOPTENKPADSLPE	-----APNEKPKVPKPSNSTDNGML	---NPEGNVGSDPMLDPALEE	967
Db	998	GTCSNPPSSQSVGLTLEFEFNLPVLFPRTKFSISVSVIEQPTNINSNEMRE	---FTEQ	1056	
QY	968	APAVDPVOE	---KLEKFTASYGLGLDSVIF	994	
Db	1057	ITALDPLSLDKQEKYNQLNLTWRHYSILF	1084		
RESULT 10					
SC16_YEAST					
ID	SC16_YEAST	STANDARD;	PRT;	2195	AA.
AC	P48415;	Q02822;			
DT	01-FEB-1996	(Rel. 33, Created)			
DT	15-JUN-2002	(Rel. 41, last sequence update)			
DT	15-JUN-2002	(Rel. 41, last annotation update)			
DE	Multidomain vesicle coat protein.				
GN	SEC16 OR YPL085W OR LPF1W.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=96017704; PubMed=7593161;				
RA	Espenshade P., Gimeno R.E., Holzmacher E., Teung P., Kaiser C.A.;				
RT	"Yeast SEC16 gene encodes a multidomain vesicle coat protein that				
RL	interacts with Sec23p."				
RL	J. Cell Biol. 131:311-324(1995).				
RP	[2]				
RC	SEQUENCE FROM N.A.				
RA	STRAIN=S288C / AB972;				
RX	MEDLINE=97313271; PubMed=9169875;				
RA	Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoerge W.,				
RA	Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,				
RA	Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,				
RA	Chung E., Church C.M., Coster F., Davis K., Davis R.W.,				
RA	Dietrich F.S., Dilius H., Dipaolo T., Dubois E., Duesterhoeft A.,				
RA	Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,				
RA	Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,				
RA	Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,				
RA	Kamp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,				
RA	Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,				
RA	Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,				
RA	Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,				
RA	Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,				
RA	Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,				
RA	Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,				
RA	Zhong W.W., Zollner A., Vo D.H., Hani J.;				
RL	"The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI."				
CC	Nature 387:103-105(1997).				
CC	-!- FUNCTION: INVOLVED IN THE BUDDING OF TRANSPORT VESICLE FROM THE				
CC	ENDOPLASMIC RETICULUM. THE C-TERMINAL INTERACTS WITH SEC23 AND				
CC	WITH THE CYTOSOLIC DOMAIN OF SED4. COULD THEREFORE BE A				
CC	CONSTITUENT OF COPII VESICLE COAT. N-TERMINAL OVEREXPRESSION				
CC	CAUSES A LETHAL SECRETION DEFECT.				
CC	-!- SUBCELLULAR LOCATION: ON THE ENDOPLASMIC RETICULUM AND ON VESICLES				
CC	WHICH BUD FROM IT.				
This SWISS-PROT entry is copyright. It is produced through a collaboration					

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U23819; AAC49088.1; -
DR EMBL; U41849; AAB68254.1; -
DR SCD; S0006006; SEC16.
KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum.
FT DOMAIN 1997 2094 LYS-RICH.
FT MUTAGEN 1059 1059 L->S; IN SEC16-4; TS ACCUMULATION OF ER
FT MUTAGEN 1084 1084 L->P; IN SEC16-3; TS ACCUMULATION OF ER
FT MUTAGEN 1089 1089 L->P; IN SEC16-2; TS ACCUMULATION OF ER
FT MUTAGEN 1231 1231 W->R; IN SEC16-1; TS ACCUMULATION OF ER
FT CONFLICT 522 522 MISSING (IN REF. 1).
FT CONFLICT 560 560 I -> F (IN REF. 1).
SQ SEQUENCE 2195 AA; 241694 MW; 757B7A7231BEE6F0 CRC64;

Query Match 3.5%; Score 185; DB 1; Length 2195;
Best Local Similarity 21.6%; Pred. No. 0.57;
Matches 223; Conservative 124; Mismatches 418; Indels 266; Gaps 54;

QY 72 GKVPYDALFSEELMKDPNQLKADIVNEVKGG-----VLIKVDGK---- 113
DB 1283 GNPV-----LSNEP--VMADSDVIFESIGNMNFISILWDEIYEYFSDPKFGF 1331
QY 114 -----YVYIKDAHADNVRTK--DEINRQ-KQEHVKDNEKN-----SNVAVARQS 157
DB 1332 SSILPKIYHASNLOGLNSLGTKYTDVLSVSRVLPKPKDILTINLTRELSVARSLS 1391
QY 158 GRVTTNDGVNPA-----DIIDTGNAYIVPHCG-----HYHYIPKSDL 197
DB 1392 ----SNTGWLAKFLSVNGQLDKSFNKYI---GGDDIDALNKKNDKKVDFGFTPCSSA 1444
QY 198 SASLAAAKAHLAKGNQPSQLSYSTASDNNTQSVAKGST--SKPANKSENLSQLLKL 255
DB 1445 NSSTVDLTOTFTFQAQVTSQSVDVDTALLHNAHNPVSHVLSKFSNVS-----KGL 1497
QY 256 YDPSAQRSESGLYFDPAKILSRTPNGVAIPH-----GDHYHPIPSKLSALEE 306
DB 1498 VEANLYTHRIGDSLOGSPQR-IHNTQFAAAEPQMASLRVRVTDQHTNEKALKSQQILEK 1556
QY 307 KIARMPISGTGSTVSTNAKPNVSWSLGSLSSNPSSLTSKELSSASDGYIFNPKDIYE 366
DB 1557 KSTAYTPQGNHSPVMEKSNVNSVLPADFAPP-----KLGTVPSNYVSSP-DLVR 1608
QY 367 ETATYIVRHGDHFHYIPKSNQIGQPTLPNNS-----LATPS-PSLEINPCTS--HEKHEE 419
DB 1609 RES---IISTGSFELPPPK---ICGPTKANSSQGLSMYSVEALPIDVPVQVHETGYN 1662
QY 420 DGYGFANRIIADDES--GFVWSHGDNHNYFFKKDLTTEQIKAAQKHLBEVTKSHGLDS 477
DB 1663 DFNKKSQKSMPEDESHSDNSADQNTLKSDADVTDMDIEGPGFNDVK---NLLPM 1719
QY 478 LSHSODYPGNAXEMKDLKKIEKTAGIMQYGVKRESIVYNKERNALYIPGHGHHAD 537
DB 1720 EPNHQPTSTVNPQ-----TISDDIQIL-QTNVEVRCTDASKMNSL--PSIENERS- 1769
QY 538 PIDEHPVGIHSHSNYELFKPEEGVAKGCKVYTGEBLTVNVLKKNSTNNQNFLLA 597
DB 1770 --SEQFENISKASS--AYLESTGSLSE-NRPLTQDE--NSIETVGOSTY-----LP 1816
QY 598 NGOKRVSFPPPELEKLGINMLVKLITPDGKVLKSVKGVGEGVGNTA----- 647
DB 1817 AG-----SISMEAKPISQVDVPRVNNKASKLYEHOHMAPKPKSTDAT 1860
QY 648 --NFELDQYLPQGTFTKYTIASKDPEVSDGTFTVPTSLAYKMASQ-----TIFYPF- 698

DB 1861 KKNY----SPYVP-----QSTAASADGDE-----STILKTSFALYARTHQAHASNPQYFFLV 1909
QY 699 -HAGDTYLRVNPQFAVPKGTDALVRVDFHGNAVLENKYKVGKELPIPKLN----- 750
DB 1910 NQANET-----ASFELSESTSQ-----QSNNGVASENRFS-----PIKKAEEVVEKDTF 1953
QY 751 QGTRTAGNK-----LPVTFMANAYL-----DNSTYIYVEVPILEKENQTKPSIL 796
DB 1954 QPTIRKASTNQYRAPKPLESDADKYNDVIEDESDDDNSTDEAKNRKEEKKNNVKKETK 2013
QY 797 POFKRNKAQENSKLDEK-----VEEPKTSKVEKEKLSGTSNSTLEEVPTVDPV 849
DB 2014 P-----SNKDIDDKSNGWEGWLLKOTGDKVYKAKLGHKNTLIYDEKLRWYKDAT 2065
QY 850 QEKVAKPAESYGMKLENVLFNMD-----GTIELYLP-----SGEVIKKNWADTGE-- 895
DB 2066 EEEKOKIIESSAPPPPIYKRRDGGPKTKPRSGPINNSLPPVHATSVIPNN--PITGEPL 2123
QY 896 -----APOG-NGENKPSNGKVS--TGTVENQPTENKPADSLPEAPNEKP--VRPENST 944
DB 2124 PIKTSPTGTPNPNNSPSPSSPISRLSGVNLTSKKNGLDLLLLAGGPKPASTRRKKKT 2183
QY 945 DNGMLNPEGNV 955
DB 2184 ARGVNVMDNI 2194
RESULT 11
SPAZ_YEAST
ID SPAZ_YEAST STANDARD; PRT; 1466 AA.
AC P23201;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SPAZ protein.
GN SPA2 OR PEAL OR YLL021W OR LI209.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288c;
RX MEDLINE-91009481; PubMed-2211820;
RA Gehring S., Snyder M.;
RT "The SPA2 gene of Saccharomyces cerevisiae is important for
pheromone-induced morphogenesis and efficient mating.";
RL J. Cell Biol. 111:1451-1464(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288c;
RA Purnelle B., Goffeau A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN PHEROMONE-INDUCED MORPHOGENESIS AND
EFFICIENT MATING, PERHAPS AS A CYTOSKELETAL PROTEIN.
CC -!- SUBCELLULAR LOCATION: LOCALIZES A SHARP PATCH AT THE SHMOO TIP
(MATING PROJECTION) WHICH IS THE SITE OF POLARIZED CELL GROWTH.
CC -!- SIMILARITY: TO PROTEINS THAT FORM COILED-COIL STRUCTURES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X53731; CA37763.1; -
DR EMBL; X97560; CA66170.1; -
DR EMBL; Z73126; CA97469.1; -
DR PIR; A36426; A36426.
DR SGD; S0003944; SPA2.

RL Biochim. Biophys. Acta 1075:248-258(1991).

CC -!- FUNCTION: May play a role in intercellular signaling and in

CC connecting cells with the extracellular matrix. May take part in

CC the regulation of cell motility, growth and differentiation. Binds

CC hyaluronic acid.

CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.

CC -!- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2

CC and V3; are produced by alternative splicing.

CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed

CC in the central nervous system, and in a number of mesenchymal and

CC epithelial tissues; the major isoform V2 is restricted to the

CC central nervous system.

CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development

CC (by similarity).

CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.

CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.

CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTOGLYCAN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@sib-sib.ch.

CC -----

CC EMBL: AF060456; AAC24358.1; -

CC EMBL: AF060457; AAC24359.1; -

CC EMBL: AF060458; AAC24360.1; -

CC EMBL: AF060459; AAC24361.1; -

CC HSP: P01132; LEPC.

CC InterPro: IPR000152; Asx_hydroxyl.

CC InterPro: IPR000561; EGF-like.

CC InterPro: IPR000742; EGF_2.

CC InterPro: IPR001881; EGF_Ca.

CC InterPro: IPR003599; Ig.

CC InterPro: IPR003006; Ig_MHC.

CC InterPro: IPR003600; Ig_Like.

CC InterPro: IPR001304; Lectin_C.

CC InterPro: IPR000538; Link.

CC InterPro: IPR000436; Sushi_SCR_CCP.

CC Pfam: PF00008; EGF; 2.

CC Pfam: PF00047; Ig; 1.

CC Pfam: PF00059; lectin_c; 1.

CC Pfam: PF00084; sushi; 1.

CC Pfam: PF00193; Xlink; 2.

CC PRINTS: P01285; LINKMODULE.

CC ProDom: PD000918; Link; 1.

CC SMART: SM00032; CCP; 1.

CC SMART: SM00034; CLECT; 1.

CC SMART: SM00181; EGF; 2.

CC SMART: SM00179; EGF_CA; 2.

CC SMART: SM00409; Ig; 1.

CC SMART: SM00410; IG_Like; 1.

CC SMART: SM00445; LINK; 2.

CC PROSITE: PS00010; ASX_HYDROXYL; 1.

CC PROSITE: PS00615; C-TYPE LECTIN_1; 1.

CC PROSITE: PS00041; C-TYPE LECTIN_2; 1.

CC PROSITE: PS00022; EGF_1; 2.

CC PROSITE: PS01186; EGF_2; 1.

CC PROSITE: PS01187; EGF_CA; 1.

CC PROSITE: PS01241; LINK; 2.

CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;

CC Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;

CC Hyaluronic acid; Alternative splicing.

FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 3381 VERSICAN CORE PROTEIN.

FT DOMAIN 37 138 IG-LIKE V-TYPE DOMAIN.

FT DOMAIN 168 245 LINK 1.

FT DOMAIN 266 347 LINK 2.

DOMAIN	349	1336	GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN)
FT	1337	3074	GAG-BETA.
FT	3074	3110	EGF-LIKE 1.
FT	3110	3148	EGF-LIKE 2, CALCIUM-BINDING.
FT	3161	3275	C-TYPE LECTIN.
FT	3280	3338	SUSHI.
FT	44	131	BY SIMILARITY.
FT	173	244	BY SIMILARITY.
FT	197	218	BY SIMILARITY.
FT	271	346	BY SIMILARITY.
FT	295	316	BY SIMILARITY.
FT	3078	3089	BY SIMILARITY.
FT	3083	3098	BY SIMILARITY.
FT	3100	3109	BY SIMILARITY.
FT	3116	3127	BY SIMILARITY.
FT	3121	3136	BY SIMILARITY.
FT	3138	3147	BY SIMILARITY.
FT	3154	3165	BY SIMILARITY.
FT	3182	3274	BY SIMILARITY.
FT	3250	3266	BY SIMILARITY.
FT	3281	3324	BY SIMILARITY.
FT	3310	3337	BY SIMILARITY.
FT	57	57	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	331	331	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	352	352	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	817	817	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	965	965	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	1017	1017	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	1333	1333	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	1393	1393	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	1437	1437	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	1463	1463	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	1653	1653	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	1974	1974	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	2045	2045	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	2074	2074	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	2103	2103	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	2263	2263	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	2290	2290	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	2356	2356	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	2623	2623	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	2641	2641	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	2919	2919	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	3052	3052	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	3354	3354	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	3364	3364	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	349	349	P -> R (IN ISOFORM V1 AND V3).
FT	350	1336	MISSING (IN ISOFORM V1).
FT	1337	3074	MISSING (IN ISOFORM V2).
FT	350	3074	MISSING (IN ISOFORM V3).
FT	25	25	MISSING (IN REF. 2).
FT	51	51	MISSING (IN REF. 2).
FT	89	89	N -> D (IN REF. 2).
FT	96	96	Q -> D (IN REF. 2).
FT	346	346	C -> R (IN REF. 2).
SQ	SEQUENCE	3381 AA; 369984 MW; F09716FA778D459 CRC64;	

Query Match 3.5%; Score 184; DB 1; Length 3381;

Best Local Similarity 20.1%; Pred. No. 1.2;

Matches 199; Conservative 135; Mismatches 340; Indels 316; Gaps 48;

Qy	233	VAKGTSKPAKSENQSLKE-LYDPSAORYSESQGL-VFDPKTIISRTPNGVAIPHG	290
Db	1649	LSQRTNTPRYTSDTSKIMITESLVDPVPTTIIISIQVSNVVPKFKVRET	1699
Qy	291	DHYFTI---PVSKLSALEEKIARVPISGTGSTVSTNA-----KPNQVSSLSGLS	338
Db	1700	DTYEWFSPPLEETTRKEEKG---TTGTASTVEVHSPTQLDQVSPSELESS-----S	1751
Qy	339	SNP---SSLTTSKELSSASDGYIENPKDIVETATATVIRHGDHFHVPKSNQIQGPTLP	395
Db	1752	ETPPDDSAATPKSFTSQ-----MTPTQSERETTSTWVF-----KETEY-----LD	1793

```

QY 396 NNSLATPSPSL-----PINTGSHKHEBDGYGFD-----NRI 429
DB 1794 NLAAQTDPDLSLSSQGVLEVPVPGSPVSLFMQSGGEAAVDPETTTVSSLSLNIETPEI 1853
QY 430 IADESSGFVNSHGDNHVFYFKKLTETQIKAAQKHLAEVTSNGLDLSLSSHEQDYPGNA 489
DB 1854 LAKEBAAGASPNVETVPFEP--TEQVLSTAVDREVAETISQTSKNLVSEISGEPTHR 1911
QY 490 KEMK--DLDKKIEEKIAGIKQYGVKRESIVYNKKNAILIYHGDHHAAD-----537
DB 1912 AEIKGFSTDFPLEEDFSDFREYST--VSPYITKEEIVVMWEGSGDAAFKDTOMLPSVTPT 1969
QY 538 -----PIDEKHPGVIGHSHSNLYELFPEEGVAKKEGKKNVYTGELTNVNLKNKTFNN 591
DB 1970 SLSNHTADSSEP---GSTLVSTAFWEPTASAG-----SGEPLLSV-----SSVD 2016
QY 592 QNFTLANGKRVSFSPPELEKKLG-----INML-----VKLITPDGKYLE- 632
DB 2017 QVFPAAAG--KASGTDSPFIDQRLGEGAINETDQRTSTILPTAEASTKASTEKEGVKEN 2074
QY 633 -----KVSCKVTEGVGNIA-----NFE-LDQYPLPGQTFKY 663
DB 2075 HTVSMDFPTEVPDELWPRQVNPVRGNSGEIVSEKTEQESFEFLQSSVAPEQT---2131
QY 664 TIASKDYPE--VSYDGTFTVPTSLAYX-----688
DB 2132 TFDSTFPEPGLQTTGYFTLTTRKTYSTDERNEDEVISLADVSTPTLDSKGLVLYTTLPE 2191
QY 689 -----MASOTIEVPPH---AGDTYLRVNPQFAPKGTDAALVRVDE---EHNAY 732
DB 2192 VTEKSHFLATASVETSPAESVIASTGIEESIKFPPKVTSPIKESDTDLIFSLGLS 2251
QY 733 LE-----NNTKVGGEIKLIPKLNQGTTR-----TACNKIPVTPMAN-----768
DB 2252 GEEVLPTLGSVNFTEIEQVLTSLPLTSQVOSLEASILNDTSGDYEGMENVANEMRPLIS 2311
QY 769 -----AYLDNQSYIIVEPILSKENOTDKPSILP--QPKRNKAQENSKLDEKVEEPPKSEKV 823
DB 2312 KDSIFDEGEVASTTLEILSDARTEGTPTAPLTFGTGQPNQONTHRAEIQIS---2368
QY 824 EKEKUSGTNSTSLTEVTPDVPQEKVAKFAESYGMKL-----ENVLFNMD 872
DB 2369 RPQPLTDQVSSNSVYATKTKTATPATDFLAR---TYDLEMAKGFVTPTPKPSDLFEVS 2425
QY 873 GTIELYLPSSGVIRKKNMADTFGEAPOGNGENKPSKNGKSTGTVENOPTENKPADSLPEA 932
DB 2426 GE-----CSGEL-----DAGAEVHAGMTQATROG--STTFVSDRSLEKH-----PKV 2467
QY 933 PNEKQVKNENSTDNM-----LNPEGNVGS-----DPMPLDPALEEAP--AVDPVQE- 976
DB 2468 PSVEAV-----TVNGFTTSMVPLPHEPREGSPETATGTPTASTASYEKATEGAADSFQDH 2522
QY 977 -----KLEKFTASYGLGLDSVIFNMD 997
DB 2523 FNGFKDSTLKPDKRKAT-----ESILIDLD 2547

RESULT 13
MSPI_PLAF3
AC P19598: 025921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_taxID=5834;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.

```

```

RX MEDLINE=88166657; PubMed=327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
RL precursor p190 of plasmodium falciparum lacks tripeptide repeats.";
RN EMBO J. 6:4137-4142(1987).
RP SEQUENCE OF 1032-1582 FROM N.A.
RA MEDLINE=95354793; PubMed=7628566;
RT Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RL merozoite surface antigen-1.";
RC Exp. Parasitol. 81:47-54(1995).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M35727; AAA29715.1; -
CC EMBL; Y00087; CAA69280.1; -
CC EMBL; Z35326; CAA84555.1; -
CC PIR; S06286; S06286.
CC InterPro: IPR000561; BGF-like.
CC Pfam: PF00008; BGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82ALEI5994BCAD6 CRC64;
Query Match 3.4%; Score 181.5; DB 1; Length 1682;
Best Local Similarity 20.1%; Pred. No. 0.57;
Matches 238; Conservative 173; Mismatches 417; Indels 359; Gaps 63;
QY 3 YALNQRSQEN----KNNRVSYVDGSSQSKSENITPDQVSKQKGIQAEQIVKTIIDG 58
DB 511 YTNVKEQRYNNKFSNNVNV---QKLKALSYLED-YSLRGI-----553
QY 59 YVTSHGDRHYHYNGKVPYDA---LFSEEL-----LMKDPNVLKADIVN- 100
DB 554 ---SEKDFNHYTLTKGLEADIKKLTETKSENKILEKNSKLTHTSANASLEVSIVKL 610
QY 101 EVKGGYII-KVDG--KYVYVLKAAHADNVT-----KDEINPQKQ- 138
DB 611 QVQVLLIKKIEDLRKIELFLKNAQLKDSHVNPNIYKPNKPPYVILVKREVDKREF 670
QY 139 ----EHWKNEKVNVA---VARSGRYTTNDGVNFPADIIETDGNAYIVPFGGHH 190
DB 671 IPKVKMLKKEQAVLSITQPLVAASE---TTEDG-----GHSTHTLSQSGETE 716
QY 191 YIPKSDLSASELAATAKHAHLAKGNQPSQLSYSTASDNNTQSVAKGTSKPANK-----244
DB 717 VTEETEETVGHVTTVTTLPPKVKVNVSEIEKHSND--NSQALTKVYLKLDLDFLTKSY 775

```

```

QY 245 -----SENLOSLLKELYD-SPSAQRYSESDGLVDFDPAKIISRTPNQVAPLPHG----- 290
Db 776 ICHKYILVNSMDQKLEVLNLTPENELKSCDRL-----DLLFNQNNIPAWSLYDS 830
QY 291 -----DIYHPTPYSK-----LSALEEK--IARVV-----PSGTGTVST-NAPKNEVWSS 333
Db 831 MNNDLQHLFELYQKEMYLYHLKKEENHIKKLEEPKQITGTSSTSSPGNTTNTVAQSA 890
QY 334 IGLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHFKPKSNQI-GQP 392
Db 891 THNSNQOQSNASS---TNTQNG-----VAVSSGPAVVEESHDLTYLSINDLKGI 940
QY 393 TLN--NSLTPSPSLPINCSTGHEHEDEGDFDANRIIAEDSGFVMSHGDHNYFFK 450
Db 941 SLNLGNKTKVPNP-LTIST-TEMEKFYENILKI-MIPIFNDDIKOFVKSNSK-----VI 992
QY 451 KDLTEEQIKAAQKHLBEVKTS-----HN-----GLDSLSSHEDQYPGNAKEMD---LD 496
Db 993 TGLTETQKNALNDEIKKLDLQSLFDLYNKYKLDRLFNKKELGQDKMOIKKLTLLK 1052
QY 497 KKEEKTAGIMKQYKRE-SIVVNKEKNAI-----YPHGDH 533
Db 1053 EQLESKLSLNNPHNLQFVSFFNKKKEAEIAETENTLENTKILLKHYKGLVYVNGE- 1111
QY 534 HHADPIDEHRPVGI--GHSHSNYELFKPEGVAKKGNKVYTGEE-----LTNVNL 583
Db 1112 --SSPLKTLSEVSIQTEDNYANLEKFRVLSKIDKLNLDLHOKKLSLSSGLHLITE 1169
QY 584 LKNSNTNNQNTLANGQKRVSFSPPELEKKLGINMLVKL---ITPDGKVLKRVSGKVF 640
Db 1170 LK-EVKNKNY-----GNSPSENKK--VNEALKSYENFLPEAKVTVTVP- 1213
QY 641 EGVGNIANPELDQPLPGOTFKYTI-----ASKDYPEVSDGTFTVPTSLAYMASQTF 595
Db 1214 -----PODVTPLSPVRSVSGSGSTKEETQIPTSG-----SLTLEQOQVOL 1256
QY 696 YPHAGDTYLRVNPQFAVPGTD-----ALVRVEDEFHGNAYLENNKVGSEIKLP 745
Db 1257 QNYDEEDSLVLPFGESEDNDEYLDQVVTGEAISVTMDNLSG--FENEVDVYLK-P 1313
QY 746 IPKLNQGTTRTAGNKP--VVFMANA-----YLD-----NQSTY 777
Db 1314 L-----AGVYRSUKQKEKIFNFTLNLDILNSRLKRRKYFLDVLDESLLMQFKHISNEY 1369
QY 778 IVEVPILEKENOTDKPSILPQFKRKN-----AOENSKLDEKV-----EEPKTSEKV 823
Db 1370 IIE-DSFKLLNSDOKNTLLKSKYIKESVENDIKPAQEGISYVEKVLAKYKODLESIKRV 1428
QY 824 EKEKLSSETNSNSTLEEVTPDVOEKVAKFA-----PSYGMKLENVLFNM-- 871
Db 1429 IKE---EKEFPSPPTTPSPAKTDEQKESKFLPPLTNIEYLNLYNNLVNKIDYDLINLKA 1485
QY 872 -----DGTTELYLPSEGVKKKNADFTGEAPOGNGENKPS 906
Db 1486 KINDCNVEKDEARVKTILKSLDKAIDDKLDLP-----KNPYDFEATKKLINDOTKRD 1537
QY 907 ENKG-VSTQTVNQPT-----ENKPADSLPEAPNE--KPYKPN 943
Db 1538 MLGKLLSTGLVQNFNPNTIISKLEIEGKFQDMLNLSHQCVKQCPQNS 1584

RESULT 14
BUD3_YEAST
ID BUD3_YEAST STANDARD: PRT: 1636 AA.
AC P25558; P25557; P87007;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bud site selection protein, BUD3.
GN BUD3 OR YCJ014W OR YCJ14W/YCJ13W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

```

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95247824; Pubmed=7730410;
RA Chant J., Mischke M., Mitchell E., Herskowitz I., Pringle J.R.;
RT "Role of Bud3p in producing the axial budding pattern of yeast.";
RL J. Cell Biol. 129:767-778(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RA Stateva L.I.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Gronadka R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP REVISIONS.
RA Mewes H.-W.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CO-ASSEMBLES WITH BUD4 AT BUD SITES. BUD4 AND BUD3 MAY
CC COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
CC DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
CC ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U17580; AAA86315.1;
DR EMBL; X59720; CAA42346.2;
DR PIR; S19341; S19341.
DR PIR; S19340; S19340.
DR SGD; S0000520; BUD3.
DR InterPro; IPR000219; RhoGEF.
DR SMART; SM00325; RhoGEF; 1.
KW Cell cycle.
SQ SEQUENCE 1636 AA; 184717 MW; 9E4E46BA5C3A3F69 CRC64;

Query Match 3.4%; Score 179.5; DB 1; Length 1636;
Best Local Similarity 18.7%; Pred. No. 0.68;
Matches 186; Conservative 135; Mismatches 327; Indels 347; Gaps 47;

QY 71 NGKVPYDALFSEELMKDPNTYQLKDADIVNEVKGGYIIKVDGKYVYVLKDAHADNVTK 130
Db 722 NSSMAKOLLVCNENLIKLEHQLEEV----- 747

QY 131 DEINPQKQEHVKDNEKVNNSNVA-----RSQRYTTNDGYVFNPAIIEDTQNA 180
Db 748 -----KHPSTDHRAVNSKLSGASDFDTEHKRSYGTITTFRSYTSDLKD----- 793

QY 181 YIVPHGGHYHVPKSDLSASELAATAKAGHLAGKNQPSOLSYSSSTASDNNTOSVAKGSTS- 239
Db 794 --SPGDSNVTKET-----KEILPVKPTKSSKSPKPKKSTSFST-KESQTA 835

QY 240 -----KPANKSENQLSKELYDPSPAQRYSESDGLVDFDPAKIISRTPNQVAPLPHGDHY 293
Db 836 AEHTEKKPKNKGKGFVGLKNVFGSKSKSPQVRV---PKKISQRHPKS----- 883

QY 294 HFIPYSKLISALEEKIARVPISTGTVSTNAKNEVVSLSGLSSPSSLTTSKELSSA 353
Db 894 ---PVKPKMTSEK-----SSPKRAVWSSPKIKKSTSFST-KESQTA 922

QY 354 -----SDGYIFNPKDIVEETATAYIVRHGDHFHFKPKSNQIQTPLPNNSLATPS 403
Db 923 KSSLRAVEFKSDLLIGKPPD-----VGNCAH----POENTRISSVVVROTKTKYSYN 968

QY 404 PSLPINFQTSHEKHEEDGYGFDANRIIAEDP-SGFVMSHGDHNYFFKCOL-TBEQI--K 459

```

```

DB 969 PQPVTENTSNEKNVP-----KADQSTKQDNISNFADVEVSASSYPKELDAETDDOLIGK 1024
QY 460 AAKQ-----HLEVKYTSNGLDLSLSHEDYPCGNAKEMKDLID-----KKI 499
DB 1025 ATNSSVHGKELPDLAEVTTA-NRVSTTSAGDQIDTQSEFLRAADVNLSDDEHRQN 1083
QY 500 EKIAG-----IMKQGVVRESIVNKEKNAILYPHGDDHHADPIDEKPKVCIHSUS 552
DB 1084 ESRVENDOLFQDFIPKHYRNQENI--NSSSN--LFPEG-----KYPQEBK--GVSNENT 1131
QY 553 NVELEKPEGVAKKGNKVYTGELITNVLLKNSTENNQ-----NETLANGOK 601
DB 1132 NISLKTNEDASTLOKLSPOASKVLTENSNEIKDT--NNSGKOAKDTKLGDDYSDKETAK 1189
QY 602 RVSFPEPE-----LEKGLINMLVKLIPDQGVLEKVSQGVGCVGNIAFELDQPY 655
DB 1190 EIT--KPKNFVEGITERKEIFPTIPLAPPASKI-----NFQKSPSY 1229
QY 656 LPOQFKYTIASKDYPEVSYDGTFTVPTSLAYKMASOTIFYPFHAGDYLRVNPQFAVPK 715
DB 1230 I--ELFQGRVRLDKHDAHYNK-----RLASQVSL-----SEGKVN----- 1265
QY 716 GTDALVRVDFEHGNAYLENKYGEI-----KLPIKLNQGTTRAGNKIPVTWMANAY 770
DB 1266 -FEEDAATINKSODAKAERTQISEVIEYEMOQPIP-----TYLPKAK 1308
QY 771 LDNQSTYIVVPILEKENQTKPSILQPKFRNKAQENSKLDERVEEPTSEKVEKEKLE 830
DB 1309 LDSSG-----LEKSD-----DKFEIEELKELKSKTGNEVDGN 1344
QY 831 TGNSTNSTLEVP-----VDPQVKVAKFAESYGMKLEN-----VLFNMD----- 872
DB 1345 NPPSSIBKIKPKPAKVFIRTPSVRIIGRUFEDT--RKYENGSPDISFTYDTHNDEPD 1402
QY 873 -GTIELYLPSEGVIKKNMADTGEAPQNGENKPKSENGKVTSTGVNQENPKNPADSLPE 931
DB 1403 KRLMELKPSQDEIPDDR--YT-----PAE-----EPTAEFPVEELPN 1439
QY 932 ANEKPVKPE--STDNGMLNPEGNGVSDP--MLD 962
DB 1440 TPRSINVTTSNKSPTDKL--SSGNIDQKTELLD 1472

RESULT 15
PGCV_RAT
ID PGCV_RAT STANDARD: PRT: 2738 AA.
AC Q9ERB4: O08592: O88564: Q9R1K4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (pg-M) (Glial
DE hyaluronate-binding protein) (GHAP) (Fragments).
GN CPBG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.
RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=93327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wright T.N.;
RA "Versican/pg-M isoforms in vascular smooth muscle cells.";
RT Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [2]
RN SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
RP STRAIN=Wistar Kyoto;
RX MEDLINE=98308094; PubMed=9642104;
RA Milev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,

```

```

RA Margolis R.K., Margolis R.U.;
RT "Differential regulation of expression of hyaluronan-binding
RT proteoglycans in developing brain: aggrecan, versican, neurocan, and
RL Biochem. Biophys. Res. Commun. 247:207-212(1998).
RN [3]
RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).
RC TISSUE=Kidney;
RX MEDLINE=96094159; PubMed=9434070;
RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RT "Proteoglycan expression in the normal rat kidney.";
RN Nephron 77:461-470(1997).
RN [4]
RP SEQUENCE OF 2535-2738 FROM N.A.
RC STRAIN=Soraguet-Dawley; TISSUE=Lung;
RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;
RT "Molecular cloning and characterization of two developmentally
RT regulated genes in rat lung.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: At least 3 isoforms; V0 (shown here), V3 and
CC Vint; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: In kidney is expressed in the papillary area,
CC but not in glomeruli.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF062402; AAC40166.1; -
CC EMBL: U75306; AAB51125.1; -
CC EMBL: AF084544; AAD48544.1; -
CC EMBL: AF072892; AAC26116.1; -
CC EMBL: AY007691; AAG16831.1; -
CC HSSP: P01132; IEPG.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000742; EGF_2.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR003599; Ig.
CC InterPro: IPR003600; Ig-like.
CC InterPro: IPR001304; Lectin_C.
CC InterPro: IPR000538; Link.
CC InterPro: IPR000436; Sushi_SCR_OCCP.
CC PRINTS: PR01265; LINKMODULE.
CC SMART: SMC0032; CCP. 1.
CC SMART: SMC0034; CLECT. 1.
CC SMART: SMC0181; EGF. 2.
CC SMART: SMC0179; EGF_CA. 2.
CC SMART: SMC0409; IG. 1.
CC SMART: SMC0410; IG_LIKE. 1.
CC SMART: SMC0445; LINK. 2.
CC PROSITE: PS00010; ASX_HYDROXYL. 1.
CC PROSITE: PS00615; C_TYPE_LLECTIN_1. 1.
CC PROSITE: PS00041; C_TYPE_LLECTIN_2. 1.
CC PROSITE: PS00022; EGF_1. 2.
CC PROSITE: PS01186; EGF_2. 1.

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:07 ; Search time 45.2509 Seconds
(without alignments)
4639.948 Million cell updates/sec

Title: US-09-471-255-55
Perfect score: 5315
Sequence: 1 CAVALNQHRSQENKNNRVS.....TELRLPSGEVKKNLSDFFIA 1019

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodeo.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5297	99.7	1039	16 Q9ANY1	Q9any1 streptococc
2	1279.5	24.1	844	2 Q9AG74	Q9ag74 streptococc
3	1256	23.6	816	2 Q9AHT9	Q9aht9 streptococc
4	1255	23.6	802	16 Q97QM8	Q97qm8 streptococc
5	1235.5	23.2	819	16 Q97QM9	Q97qm9 streptococc
6	1233	23.2	839	16 Q9ANY2	Q9any2 streptococc
7	1232.5	23.2	819	2 Q9ANY3	Q9any3 streptococc
8	902	17.0	825	16 Q99XV4	Q99xv4 streptococc
9	899.5	15.9	822	2 Q92HG7	Q92hg7 streptococc
10	897	16.9	825	2 Q93GT5	Q93gt5 streptococc
11	661.5	12.4	289	2 Q9AE21	Q9ae21 streptococc
12	292.5	5.5	792	16 Q99Z76	Q99z76 streptococc
13	238.5	4.5	2004	16 Q97QP7	Q97qp7 streptococc
14	235.5	4.4	1078	5 Q963T1	Q963t1 plasmodium
15	228	4.3	2485	5 Q96134	Q96134 plasmodium
16	226.5	4.3	1139	5 Q97237	Q97237 plasmodium

17	223	4.2	1964	2 Q59947	Q59947 streptococc
18	218	4.1	1873	2 Q924N7	Q924n7 enterococcu
19	216	4.1	5458	5 Q9U459	Q9u459 plasmodium
20	214	4.0	1881	16 Q8RGK2	Q8rgk2 fusobacteri
21	214	4.0	1939	5 Q25662	Q25662 plasmodium
22	212.5	4.0	1733	2 Q9K114	Q9k114 staphylococ
23	210	4.0	1694	5 Q9NHX1	Q9nhx1 plasmodium
24	209	3.9	2276	2 Q93TY6	Q93ty6 staphylococ
25	207.5	3.9	1694	5 Q9TZT5	Q9tzt5 plasmodium
26	206	3.9	1593	5 Q8X82	Q8x82 crosophilla
27	204	3.8	1268	16 Q93M90	Q93m90 clostridium
28	204	3.8	1524	10 Q8RYN2	Q8ryn2 oryza sativ
29	204	3.8	2647	5 Q9U4X0	Q9u4x0 plasmodium
30	203.5	3.8	1704	5 Q9TZT4	Q9tzt4 plasmodium
31	199	3.7	1720	5 Q25922	Q25922 plasmodium
32	198	3.7	1179	2 Q9AHK5	Q9ahk5 borrelia bu
33	196.5	3.7	881	2 Q9RFV4	Q9rfj4 streptococc
34	196	3.7	1271	5 Q58660	Q58660 plasmodium
35	195.5	3.7	1927	2 Q54875	Q54875 streptococc
36	195.5	3.7	2910	10 Q9FND5	Q9fnd5 arabidopsis
37	194.5	3.7	2806	16 Q8RI19	Q8ri19 fusobacteri
38	193	3.6	3724	5 Q77320	Q77320 plasmodium
39	192	3.6	5005	16 Q9PP25	Q9pp25 ureaplasma
40	190.5	3.6	1395	2 Q9A1S0	Q9ais0 staphylococ
41	190.5	3.6	4667	5 Q9TV19	Q9tv19 caenorhabdi
42	190	3.6	1698	2 Q9LC00	Q9lc00 staphylococ
43	189.5	3.6	3257	5 Q9V736	Q9v736 drosophila
44	189	3.6	1038	16 Q99RD2	Q99rd2 staphylococ
45	189	3.6	1236	5 Q9GTX2	Q9gtx2 plasmodium

ALIGNMENTS

RESULT 1

Q9ANY1	PRELIMINARY;	PRT; 1039 AA.
ID Q9ANY1		
AC Q9ANY1		
DT 01-JUN-2001 (TRENBLrel. 17, Created)		
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)		
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)		
DE Pneumococcal histidine triad protein E precursor (Hypothetical protein SPI004).		
DE PHE OR SP1004.		
GN Streptococcus pneumoniae.		
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;		
OC Streptococcaceae; Streptococcus.		
OX NCBI_TaxID=1313;		
RN [1]		
RP MEDLINE=21101045; PubMed=11159990;		
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,		
RA Dormitzer M., Degan R., Brewan Y.A., Barren P., Lathigra R.,		
RA Langermann S., Koenig S., Johnson S.;		
RT "Identification and characterization of a novel family of pneumococcal proteins (the Pht family) that are protective against sepsis.";		
RL Infect. Immun. 69:949-958(2001).		
RN [2]		
RP SEQUENCE FROM N.A.		
RA MEDLINE=21357209; PubMed=11463916;		
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,		
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,		
RA Durkin A.S., Whittam M., Kolonay J.F., Nelson W.C., Peterson J.D.,		
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,		
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,		
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,		
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,		
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;		
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.";		
RL Science 293:498-506(2001)		
DR EMBL; AF318956; AAR06761.1; -		

DR EMBL: AE007403; AAK75121.1;
 DR TIGR: SP1004;
 KW Signal; Hypothetical protein; Complete proteome.
 FT SIGNAL 1 29 POTENTIAL.
 SQ SEQUENCE 1039 AA: 114631 MW: 8145633FC806625C4 CRC64;

Query Match 99.7%: Score 5297; DB 16; Length 1039;
 Best Local Similarity 99.7%: Pred. No. 1.1e-241;
 Matches 1016; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 CAYALNQHRSQENKNNRVSYVDSQSQSKSENLTPOVSKQEGIAQIVIKITDOGVY 60
 DB 21 CAYALNQHRSQENKNNRVSYVDSQSQSKSENLTPOVSKQEGIAQIVIKITDOGVY 80

QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVDGKYYVYLKD 120
 DB 81 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVDGKYYVYLKD 140

QY 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVNAVARSQGRYTTNDGVYFNPADIIEDTGA 180
 DB 141 AAHADNVRTKDEINRQKQEHVKDNEKYNVNAVARSQGRYTTNDGVYFNPADIIEDTGA 200

QY 181 YTVPGHGHYHYIPKSDLSASLAALAAKAGKKNQPSQSYSTASDNNQTSVAKGSTK 240
 DB 201 YTVPGHGHYHYIPKSDLSASLAALAAKAGKKNQPSQSYSTASDNNQTSVAKGSTK 260

QY 241 PANKSENLOSLLKELYDPSQRYSESDGLVDPKAIISRTNGVAIPHGDIHYHFTPSK 300
 DB 261 PANKSENLOSLLKELYDPSQRYSESDGLVDPKAIISRTNGVAIPHGDIHYHFTPSK 320

QY 301 LSALAEKIAKRVPIISGTSTYSTNAKNEVSSLSGSSNPSSLTTSKELSSASDGYIFN 360
 DB 321 LSALAEKIAKRVPIISGTSTYSTNAKNEVSSLSGSSNPSSLTTSKELSSASDGYIFN 380

QY 361 PKDIVEETATYIVRHGDHYHYIPKSNQIQOPTLPNNSLATPSPSPINPGTSHKHEED 420
 DB 381 PKDIVEETATYIVRHGDHYHYIPKSNQIQOPTLPNNSLATPSPSPINPGTSHKHEED 440

QY 421 GYGFDANRIAEDESGFVMSGCDNNHYFFFKDLTEEOIKAAQKHLKVEKTSNGLDLSLS 480
 DB 441 GYGFDANRIAEDESGFVMSGCDNNHYFFFKDLTEEOIKAAQKHLKVEKTSNGLDLSLS 500

QY 481 HQDTPGNKAKEMKLDKIEKIAIGIKQYGVKPRESTVWKNENAIYIPHGDIHHADPID 540
 DB 501 HQDTPGNKAKEMKLDKIEKIAIGIKQYGVKPRESTVWKNENAIYIPHGDIHHADPID 560

QY 541 EHKPVGIGSHSNVELFKPEGVAKKGNKYITGEELTNVLLKNSFTNNQFTLANGQ 600
 DB 561 EHKPVGIGSHSNVELFKPEGVAKKGNKYITGEELTNVLLKNSFTNNQFTLANGQ 620

QY 601 KRVSFPFPELEKLGINKMLVLTIPDGKVLKGVKGVGEGVGNIANFELDOPYLPQGT 660
 DB 621 KRVSFPFPELEKLGINKMLVLTIPDGKVLKGVKGVGEGVGNIANFELDOPYLPQGT 680

QY 661 FKYTIAKDYPEVSDGTFTVPTSLAYKMASQITFYFPHAGDTYLRVNPQFAYPKGTDAL 720
 DB 681 FKYTIAKDYPEVSDGTFTVPTSLAYKMASQITFYFPHAGDTYLRVNPQFAYPKGTDAL 740

QY 721 VRVDFEFGNAYLENNYKVGKIKLIPKLNQGTTRAGNIPVTFMANAYLDNQSYIYE 780
 DB 741 VRVDFEFGNAYLENNYKVGKIKLIPKLNQGTTRAGNIPVTFMANAYLDNQSYIYE 800

QY 781 VPILKENOTDKPSILPOFKRKAQENSLKDEKVEEPTSEKVEKEKLSGTGNSNTSL 840
 DB 801 VPILKENOTDKPSILPOFKRKAQENSLKDEKVEEPTSEKVEKEKLSGTGNSNTSL 860

QY 841 BEVPTVDVQKFAKFAESYGMKLENVLFNMDGTIELYLPSEGVYIKKNADFTGEAPQGN 900
 DB 861 BEVPTVDVQKFAKFAESYGMKLENVLFNMDGTIELYLPSEGVYIKKNADFTGEAPQGN 920

QY 901 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPNENGVSDPM 960
 DB 921 GENKPSNGKYSTGTVENQPTENKPADSLPEAPNEKPVKPNSTDNGLNPNENGVSDPM 980

QY 961 LDPALFAPAVDPVOEKLEKFTASYGLGLOSVIFNMDGTIELRPSGEVIKKNLSDFIA 1019
 DB 981 LDPALFAPAVDPVOEKLEKFTASYGLGLOSVIFNMDGTIELRPSGEVIKKNLSOLIA 1039

RESULT 2

Q9AG74 PRELIMINARY; PRT: 844 AA.
 AC Q9AG74;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PHA.
 GN PHA.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 RN NCBI_TaxID=1313;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21246685; PubMed=11349048;
 RA Zhang Y., Masi A.W., Barniak V., Mountzouros K., Hostetter M.K.,
 RA Green B.A.;
 RT "Recombinant PhA Protein, a Unique Histidine Motif-Containing Protein
 from Streptococcus pneumoniae, Protects Mice against Intranasal
 RT pneumococcal Challenge.";
 RL Infect. Immun. 69:3827-3836(2001).
 DR EMBL: AF340221; AAK26629.1;
 SQ SEQUENCE 844 AA: 94769 MW: D738A5529FF8902 CRC64;

Query Match 24.1%: Score 1279.5; DB 2; Length 844;
 Best Local Similarity 40.5%: Pred. No. 2e-52;
 Matches 301; Conservative 101; Mismatches 182; Indels 159; Gaps 21;

QY 1 CAYALNQHRSQENKNNRVSYVDSQSQSKSENLTPOVSKQEGIAQIVIKITDOGVY 60
 DB 20 CSEYGLGYQAARTKENVNRSYIDGKQATQNTENLTPEVSKREGINAEQIVIKITDOGVY 79

QY 61 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVDGKYYVYLKD 120
 DB 80 TSHGDHYHYNGKVPYDALFSEELMKDPNQLKADIVNEVKGYYIKVDGKYYVYLKD 139

QY 121 AAHADNVRTKDEINRQKQEHVKDNEKYNVNAVARSQGRYTTNDGVYFNPADIIED 176
 DB 140 AAHADNVRTKDEINRQKQEHVKDNEKYNVNAVARSQGRYTTNDGVYFNPADIIED 199

QY 177 TGNAYIVPHGHHYHYIPKSDLSASLAALAAKAGKKNQPSQSYSTASDNN 228
 DB 200 TGNAYIVPHGHHYHYIPKSDLSASLAALAAKAGKKNQPSQSYSTASDNN 259

QY 229 -----NTQSAKSTKPAKSENLOSLLKELYDPSQRYSESDGLVDPKAIISR 280
 DB 260 SVSNPGTNTNTNSNTNSQASQSDIDSLKQLKPLSQSHVESDGLIFDPAQITSR 319

QY 281 TPNGVAIPHGDIHYHYIPKSDLSASLAALAAKAGKKNQPSQSYSTASDNN 328
 DB 320 TANGVAIPHGDIHYHYIPKSDLSASLAALAAKAGKKNQPSQSYSTASDNN 379

QY 329 --EVVSLGSLSSNPSSLTTSKE--LSSASDGYIFN-----PKDIVEETAT---AVI 373
 DB 380 SPQAPNPQAPSPNPIDKLVKAVKGVGVDGYVFEENGVPRIYPAKDLAETAAGIDSKL 439

QY 374 VRHGDHYHY-----IPKSN-----QIGOPTLPNNSLATPSPSL----- 406
 DB 440 AKQESLSHKLGAKKATDLPSSDREFFYNKAYOLLARIHQDLLDKNGQVDFEALDNLERLK 499

QY 407 -----PI-----NPGTSHEKHE-----EDGYGFDAN 427
 DB 500 DVSSDKVKLVDDILAFAPIRHPERLGKNAQITYTDDIEIQVAKLAGKYTTEDGYIFDPR 559

QY 428 RIIDEDESGFVMSGCDNNHYFFFKDLTEEOIKAAQKHLKVEKTSNGLDLSHESHEQDYS 487

```

Db 560 DITSDEGAIVTPTHSHWIKKDSLEAERAAQAAYAKE-----KGLTTPSTDHOD-SG 613
Qy 488 NAKEMKDLDR-----KIEEKIAGIMKQGVKRESIVWAKENKNAIYIPGDDHHADPIDEH 542
Db 614 NT-EAKGAELYNRRAAKVPLDRMPYNIO---YTVVEVKNGSLIIIEYDHYH-----662
Qy 543 KPVGIGHSHSNVFLFPDEGVAKKGNKYVYGEETNVNVL-----KNSTFNQNN 593
Db 663 -----NTKFFWF--DEGL--YEAPKGYSLIEDLATYKYVVEHPNRPHSNGFGNAS 710
Qy 594 FTL---ANGOKRVSFSPFPELEK 613
Db 711 DHVQRNKGADINQTEKNEEK 733

RESULT 3
Q9AHT9
AC Q9AHT9 PRELIMINARY; PRT; 816 AA.
AD Q9AHT9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DE Pneumococcal histidine triad A protein.
GN PHTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RA MEDLINE=21116976; PubMed=11179332;
RA Wizenmann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
RA Langemann S., Johnson S., Koenig S.;
RT Use of a Whole Genome Approach To Identify Vaccine Molecules
RT Affording Protection Against Streptococcus pneumoniae Infection.";
RL Infect. Immun. 69:1593-1598(2001).
DR EMBL; AF291695; AAK19155.1;
SQ SEQUENCE 816 AA; 91519 MW; 5359126A611D27ED CRC64;

Query Match 23.6%; Score 1256; DB 2; Length 816;
Best Local Similarity 41.3%; Pred. No. 2.4e-51;
Matches 299; Conservative 95; Mismatches 180; Indels 150; Gaps 19;

Qy 1 CAVALNQHRSQNKNNRVSYVDGSSQSKSENTLPDQVSKQEGIAEQIVKITDQGV 60
Db 20 CSVELGLYQARTVKNRNVSYIDGQATQKTENTLPDEVSKREGINAEQIVKITDQGV 79
Qy 61 TSHGDHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEVKGXIIVDGRYYVYLKD 120
Db 80 TSHGDHYHYNGKVPYDAIIEELLMKDPNKLKDEIVNEVKGVIKVDGRYYVYLKD 139
Qy 121 AAHADNVRTKDEINRQKHVKDNE---KYNVNVARSGRYTNDGVVFNPAIDIED 176
Db 140 AAHADNVRTKEINRQKHQSHREGGTPRNDGVALARSQGRYTTDDGVFNASDIED 199
Qy 177 TGNAYIVPHGGHYHYIPKSLDLSASELAAKAHLAKGNQFSQLSYSTASDN-----228
Db 200 TGDAYIVPHGDHYHYIPKNELLSASELAAAEFLSGRNLNSRTRYRQNSDNTSRTNVWP 259
Qy 229 -----NTQSVAGSTSKPANKSENLSQLLKELYDPSAQRYSSDGLVDPDAKLISR 280
Db 260 SVSNPGTTNTNSNNTSNQASQSDIDSLIKOLYKPLSQSHVESDGLVDPDAQITSR 319
Qy 281 TPNGVAIPHGDRHYHPIPYKSLSALEEKIARVP-----313
Db 320 TARGVAVPHGDHYHPIPYQSNSELEERARIIPLYRBSNHWVDSRPEQSPQPTPEPSP 379
Qy 314 -----ISTGTVGVSTNAK-----PNEVYSSLSGLSSNPS 342
Db 380 GPQPAPNLKIDNSLSVSQLRVKGVGEYVFEKGISRYVFAKDLPSFTVKRLSKLSKOE 439

```

```

QY 343 S-----LTTSKELSSASGQYIFNPK-DIVEETATAYIVRHG--DHFHYIPK-----SNOI 389
DB 440 VSVHTLTAKKNVAVRPQEQFYDKAYNLLTFAHKALFKNKGRNSDFOALDKILLERLNDEST 499
QY 390 GQPTLPNNSIATPSP-SLPINPGTSHKHE-----EDQYGFANDRIIA 431
DB 500 NKEILVDDLLAFAPITHPERLGRPNQSIYETEDEVRIALQADKYTTSQGYIFDEHDIIS 559
QY 432 EDSGFVMSHGDNHNYFFKCOLTEEQTKAAQKHLEE-----VKTSHNGLDLSLS 480
DB 560 DEGDAYVTPPHNGHSHGWICKDSLSDEKVAQAQYTKKGIILPPSPDADVRANPTG-DSAAA 618
QY 481 HEQDYPGNAKEMKDLKKIEKIAIGIKQYGVKRESIVNVKKNATIIYEGDHHHADP-- 538
DB 619 IYNRVKG-----EKRIPLVLPYMW--EHTVEVKNGLLI-PKHDIYHNKIFA 663
QY 539 -IDEH---KPVGIG-----HSHSNVELFKPEGVAKKGNKVYTGELITNVNLLKNSTFN 590
DB 664 WFDDBTKAPNGYLEDLFATIKYVYVHPDRPHSNDG----WGNASEHVLGKKDKHSDP 719
QY 591 NONF 594
DB 720 KNKF 723

RESULT 4
ID Q97QM8 PRELIMINARY: PRT; 802 AA.
AC Q97QM8:
DT 01-OCT-2001 (TEMBLrel. 18, Created)
DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Conserved domain protein.
GN Sp1175.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463915;
RA Tetzelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayan L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfe E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angluoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001)
DR EMBL: AE007418; AAK75284.1; -.
DR TIGR: Sp1175; -.
KW Complete proteome.
SQ SEQUENCE 802 AA; 90080 MW; 4E5CB8364EEA1833 CRC64;

Query Match 23.6%; Score 1255; DB 16; Length 802;
Best Local Similarity 41.3%; Pred No. 2.7e-51;
Matches 299; Conservative 95; Mismatches 180; Indels 150; Gaps

QY 1 CAYALNHRSGENKNNRNSYVDGSSQSKSENLPDQVSQKEGIAEQIVIKITDQGY 60
DB 6 CSYELGYQARTYKNNRNSYVDGKQATQKTENLPDEYSKREGINAEQIVIKITDQGY 65
QY 61 TSHGDHYHYNGKVPYDALFSPELLMKDPNKLQADIVNEVKGGYIIVKDGKYYVYLKD 120
DB 66 TSHGDHYHYNGKVPYDAIISPELLMKDPNKLKDEDIVNEVKGGYIIVKDGKYYVYLKD 125
QY 121 AAHADNVRTKDEINRQKEHVKNDE----KVNNSVAVARSQGRYTTNDGYVFNPAIDIT 176

```

Query Match	23.6%	Score 1255;	DB 16;	Length 802;
Best Local Similarity	41.3%	Pred. No. 2.7e-51;		
Matches 299;	Conservative	95;	Mismatches 180;	Indels 150;
				Gaps 19;

[illegible]

Db 126 AAHADNVRTKEEINROKOEHSQHREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIIED 185
 QY 177 TGNAYIVPHGGHYHYIPKSDLSASLAAAHLAGKNQPSQLSYSTASDN-----228
 Db 186 TGDAYIVPHGGHYHYIPKNEISASLAAAFULSGRNLNSRTYRQNSDNTSRNWP 245
 QY 229 -----NTOSVAKGSTPKANKSENQLSKELXDPSAQRYSESGLVFPDPAKILSR 280
 Db 246 SVSNPGTNTNTSNNSNTNSOASQNDISLLKOLYKPLSQRHVESDGLVFPDAQITSR 305
 QY 281 TPNGVAIPHGGHYHYIPYKLSALEEKTARMVP-----313
 Db 306 TARGVAVPHGGHYHYIPYKNSLELEERARIPLRYRSNHWVDSRPEQSPQPTPEPSP 365
 QY 314 -----ISGTGSTVSTNAK-----PNEVSSLSGLSSNPS 342
 Db 366 GPQAPNLKIDNSLSVQSVKRYGEGYVFEKGISRYFAKDLPSFVKNLESLSKOE 425
 QY 343 S-----LTSKLSASDGYIFNPK-DIVEETATAYIVRHG--DHFHYIPK-----SNQI 389
 Db 426 SVSHTLTAKKNVAPROGFEYDKAYNLLTAHKAIPENKGRNSDPQALDKLLERLNDEST 485
 QY 390 GQPTLPNNSLATSP-SLIPNGTSHKHE-----EDGYGFDAHRIIA 431
 Db 486 NKEKLVDDLLAFAPITHPERLGRPNISQIEYTEDVRIAGLADKYTSDGYIFDEHDIIS 545
 QY 432 EDESGFVMSHGDHNYFPKDKLTPEQIKAAKHLKE-----VKTSNGLDLSLS 480
 Db 546 DEGDAYIVPHGSHWICKDLSKDKVAQAAYTKKGIPLPPSPDADYKANPTG-DSAAA 604
 QY 481 HQDYPGNNAKEMKLDKIEKIGIKWQYKRESIVVKNKNAIYPHGDHHHADP---538
 Db 605 IYNRVKG-----EKRIPLVRLPMV--EHTVEVKGNLII-PKHGDYHNKIFA 649
 QY 539 -IDER---KPVGIG---HSHSNVELEKPEGVAKKRGKGVYTGELTNVNNLLKNSTEN 590
 Db 650 WEDDHTYKAPNGYTLDFATIKYIVVEHPDERPHSNDG-----WGNASEHVLGKDKHSEDP 705
 QY 591 NONF 594
 Db 706 NKNF 709
 RESULT 5
 Q97QM9 PRELIMINARY; PRT: 819 AA.
 AC Q97QM9;
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Conserved domain protein.
 GN SP1174.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group: Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 RN NCBI_taxid=1313;
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Heidelberg J., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Desoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Felblyum T.V., Angiuoli S., Dickson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Wang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae";
 RL Science 293:498-505(2001).
 DR EMBL; AE007418; AAK75283.1; .
 DR TIGR; SP1174; .

RW Complete proteome.
 SQ SEQUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;
 Query Match 23.2%; Score 1235.5; DB 16; Length 819;
 Best Local Similarity 41.0%; Pred. No. 2.3e-50;
 Matches 297; Conservative 102; Mismatches 183; Indels 143; Gaps 24;
 QY 1 CAYALNQRS--QENKNNRYSYVDGSSQSKSENLPDQVQSKREGIQAEQIVIKITDQY 59
 Db 20 CSYELGRYQAGQDKKESNRVAYIDGQAGQAKENLPIDEVSKREGINAEQIVIKITDQY 79
 QY 60 VTSBGDHYHYNGVVPYDALFSEELLKMDPNYOLKQADIVNEVKGYIIKVDGKYVYLK 119
 Db 80 VTSBGDHYHYNGVVPYDALFSEELLKMDPNYOLKQADIVNEVKGYIIKVDGKYVYLK 139
 QY 120 DAAHADNVRTKDEINROKOEHRVD-NEKVNNSVAVARSQGRYTTDDGYIFNASDIIEDTG 178
 Db 140 DAAHADNVRTKEELKQKQERSHNHNSRADNAVAARAQGRYTTDDGYIFNASDIIEDTG 199
 QY 179 NAYIVPHGGHYHYTPKSDLSASLAAAHLAGKNQPSQLSYSTASDNNTQ---SVAK 235
 Db 200 DAYIVPHGGHYHYTPKNEISASLAAAFULSGRNLNSRTYRQNSDNTSRNWP 257
 QY 236 GSTSKPA---NKENLQSLKELYDPSAQRYSESGLVFPDPAKILSRTPNGVAYPHGH 292
 Db 258 NLVTPTVYHQNGENISLLRELYAKPLSERHVESDGLIFDPAQITSRARGVAVPHGNH 317
 QY 293 YHFIPYKLSALEEKIARMVPISGTGSTVSTNAKPN-----EVVSSLSGLSSNPSL 344
 Db 318 YHFIPYQMSLEARIARIPLRYRSNHWVDSRPEPSPQPTPEPSPQAPSNPIDE 377
 QY 345 TTSKE-LSSASDGYIFN-----PKDIVEETAT---AYIVRHGDHPHY-----IP 384
 Db 378 KLVEARVKGVDGYVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESLSHKLGTGKTDLIP 437
 QY 385 KSN-----QIGQPTLPNNSLATSPSL-----406
 Db 438 SSOREFYKAYVDLLARIHQDLNKGQVDFEALDMLLERKLDVSSDKVKVLEDLLAFIA 497
 QY 407 PI-----NPGTSHKHE-----EDGYGFDAHRIIAEDESQVFNHGDH 445
 Db 498 PIHPERLGRPNAGITVTDDEIQVAKLAGKYTTEDGYIFDPDITSDGDAYVTPHMTHS 557
 QY 446 HYFVKKLTDEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYPGNKAKEMKDLK-----KTE 500
 Db 558 HWIKDLSLEAERAAQAYAKE-----KGUTPPSTDHQD--SGNT-EAKGAEAIYNRVKAA 610
 QY 501 EKIAGIKWQYKRESIVVKNKNAIYPHGDHHHADPIDEHKPVGIGISHSNYELFKPE 560
 Db 611 KKVPLDRMPYNLQ---VTVEVKNGSLIIPHYDHYH-----NIKFEWF--D 650
 QY 561 EGVAKKRGKGVYTGELTNVNNLL-----KNSTFNNQFTL---ANGKRVSEFSP 608
 Db 651 EGL--YEAPKGYTLEDLATVYKIVVEHPDERPHSNDGFGNASDHVORNGKQADINQTEK 708
 QY 609 PELEK 613
 Db 709 PSEK 713
 RESULT 6
 Q9ANY2 PRELIMINARY; PRT: 839 AA.
 ID Q9ANY2
 AC Q9ANY2;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
 DE Pneumococcal histidine triad protein D precursor (Hypothetical protein
 DE SP1003) (Fragment).
 GN PFTD OR SP1003.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group: Lactobacillales;
 OC Streptococcaceae; Streptococcus.

OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21101045; PubMed=11159990;
 RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
 RA Dormitzer M., Dagan R., Brewar Y.A., Barren P., Lathigra R.,
 RA Langermann S., Koenig S., Johnson S.,
 RA "Identification and characterization of a novel family of pneumococcal
 RT proteins (the Pht family) that are protective against sepsis.",
 RL Infect. Immun. 69:949-958(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L., E.K.,
 RA McDonald L.A., Feldblyum T.V., Angluoli S., Dickinson T., Hickey J.C.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RA "Complete genome sequence of a virulent isolate of *Streptococcus*
 RT *pneumoniae*,"
 RL Science 293:498-506(2001).
 DR EMBL: AF318955; AAK06760.1;
 DR EMBL: AE007403; AAK75120.1;
 DR TIGR: SP1003;
 KW Signal; Hypothetical protein; Complete proteome.
 FT SIGNAL
 FT NON_TER 839 839 POTENTIAL.
 SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;
 Query Match 23.2%; Score 1232; DB 16; Length 839;
 Best Local Similarity 42.8%; Pred. No. 3.1e-50;
 Matches 295; Conservative 90; Mismatches 164; Indels 140; Gaps 25;
 QY 1 CAYALNQHRS-QENKDNRRVYVDSQSQSKSENLTDPQVSKQEGIAEQIVIKITDQGY 59
 Db 20 CSYELGRHQAGQVKESNRVYIDGDAQQAENLTDPDEVSKREGINAQIVIKITDQGY 79
 QY 60 VTSHGDHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEKGGYTIKVDGKYYVYLK 119
 Db 80 VTSHGDHYHYNGKVPYDAIIEELLMKDPNQLKADIVNEKGGYTIKVDGKYYVYLK 139
 QY 120 DAAHADNRTKDEINRQKQEHVKD--VAFAESQGRYTTNDGYVFNPAIDIED 176
 Db 140 DAAHADNRTKDEINRQKQEH--SHNNGGSNDQAVVAARAQGRYTTDDGYIFNASDIID 198
 QY 177 TGNAYIVPHGHHYHYIPKSDLSASELAALAKHLAGKNMQPSQSYSTASDNNITQ---SV 233
 Db 199 TGDAYIVPHGHHYHYIPKSNLSASELAALAEAYWNGK--QGSRPSSSSSYNANPAQPRLSE 256
 QY 234 AGSTSKPA---NKSENLOSLLKELYDSPAQRYSESDGLVDFDPAKIISRTNGVAIPHG 290
 Db 257 NNLTVPTVYHQNGENISSLLRELKPLSERHVESDGLIFDPQAITSRTARGVAVPHG 316
 QY 291 DHYFTIPYKLSALEEKIARMVPISTGTSTVSTNAKPN-----EVVSSLSGS 336
 Db 317 NHYFTIPYQMSLEKRIARIIPLYRSNHWVDSRPEQSPQSTPEPSPQAPNPQP 376
 QY 337 LSNPSSLTSKE-LSSASDGYFN-----PKDIVEETAT---AVIVRHGDHFFH- 382
 Db 377 APSNPIDEKLVKAEVRKVGDFEENGYSRYIPAKDLSAETAAGIDSKLAKOESLSHKL 436
 QY 383 -----IPKSN-----QIGQPTLPN-----NSLA-----TPSPSL--- 406
 Db 437 GAKKTDLPSSDREFYFNKAYDILLARHQDLNKNKQVDFEALDNLRLERLKVPSDKVLV 496
 QY 407 -----PI-----NPGTSKEHE-----EDGYCFDANRIIAEDSGF 437
 Db 497 DDILAPLIRHPERLGRKPNQAITTDDIEIQVAKLAGKYTTEDGYIFDPRDITSGDAY 556

QY 438 VMSHGDHNYFFKKDLTBEQIKAAQKHLEEVKTSHNGLDSSSHSEQDYFGNAKEMKDLK 497
 Db 557 VTPHMTSHWIKKDSLSAEARAAQAYAKE-----KGLTPPSTHQD-SGNT-EAKGAE 609
 QY 498 -----KIEKIAIGIMKQYGVKRESIVNKEKNALIIYPHGDHHDHPIDEHKPVGISHSHS 552
 Db 610 IYNRVKAARKVPLDRMPYNLQ---YTVKNGSLIIPHYDVH-----NI 651
 QY 553 NYELPKPEGVAKKEGKNVYTGEEELTNV 581
 Db 652 KFEWF--DEGL--YEAPKGYTTLELLATV 676

RESULT 7

Q9ANY3
 ID Q9ANY3 PRELIMINARY; PRT; 819 AA.
 AC Q9ANY3.
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE Pneumococcal histidine triad protein B precursor (Fragment).
 GN PHTB.
 OS *Streptococcus pneumoniae*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OC NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21101045; PubMed=11159990;
 RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
 RA Dormitzer M., Dagan R., Brewar Y.A., Barren P., Lathigra R.,
 RA Langermann S., Koenig S., Johnson S.,
 RA "Identification and characterization of a novel family of pneumococcal
 RT proteins (the Pht family) that are protective against sepsis.",
 RL Infect. Immun. 69:949-958(2001).
 DR EMBL: AF318954; AAK06759.1;
 KW Signal.
 FT SIGNAL 1 29 POTENTIAL.
 FT NON_TER 819 819
 SQ SEQUENCE 819 AA; 92108 MW; E602CFC16CC28A5F CRC64;

Query Match 23.2%; Score 1232.5; DB 2; Length 819;
 Best Local Similarity 41.0%; Pred. No. 3.1e-50;
 Matches 297; Conservative 101; Mismatches 184; Indels 143; Gaps 24;

QY 1 CAYALNQHRS-QENKDNRRVYVDSQSQSKSENLTDPQVSKQEGIAEQIVIKITDQGY 59
 Db 20 CSYELGRYQAGQDKESNRVYIDGDAQQAENLTDPDEVSKREGINAQIVIKITDQGY 79
 QY 60 VTSHGDHYHYNGKVPYDALFSEELLMKDPNQLKADIVNEKGGYTIKVDGKYYVYLK 119
 Db 80 VTSHGDHYHYNGKVPYDAIIEELLMKDPNQLKADIVNEKGGYTIKVDGKYYVYLK 139
 QY 120 DAAHADNRTKDEINRQKQEHVKD--NEKVSNAVARSQGRYTTNDGYVFNPAIDIEDTG 178
 Db 140 DAAHADNRTKDEINRQKQESHNSRADNAVAARAQGRYTTDDGYIFNASDIIDTG 199
 QY 179 NAYIVPHGHHYHYIPKSDLSASELAALAKHLAGKNMQPSQSYSTASDNNITQ---SVAK 235
 Db 200 DAYIVPHGHHYHYIPKSNLSASELAALAEAYWNGK--QGSRPSSSSSYNANPAQPRLSNH 257
 QY 236 GSTSKPA---NKSENLOSLLKELYDSPAQRYSESDGLVDFDPAKIISRTNGVAIPHGDH 292
 Db 258 NLTVPVTHQNGENISSLLRELKPLSERHVESDGLIFDPQAITSRTARGVAVPHGNH 317
 QY 293 YHFTIPYKLSALEEKIARMVPISTGTSTVSTNAKPN-----EVVSSLSGSNPSL 344
 Db 318 YHFTIPYQMSLEKRIARIIPLYRSNHWVDSRPEPSPQSTPEPSPQAPNPQP 377
 QY 345 TTSKE-LSSASDGYFN-----PKDIVEETAT---AVIVRHGDHFFH-----IP 384
 Db 378 KLVKAEVRKVGDFEENGYSRYIPAKDLSAETAAGIDSKLAKOESLSHKLTKTKDLP 437

DR EMBL; AF062533; AAD13797.1; -
KW Hypothetical protein.
SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;
Query Match 16.9%; Score 899.5; DB 2; Length 822;
Best Local Similarity 27.0%; Pred. No. 1.6e-34;
Matches 262; Conservative 144; Mismatches 297; Indels 267; Gaps 31;
QY 2 AYALNQRSCQENKNNRVSYVDGSSQSKS--ENLTPDQVSQKEGIAQAEQIVIKITDQGY 59
Db 22 SYQLGKHGMLATKNDQIAYIDDSKGRAPKTNKMDQISAEGBISAEQIVVITDQGY 81
QY 60 VTSBGDHYHYNGKVPYDALFSEELLMDPNYQDKADIVNEKGGYIIKVDGKYVYVLK 119
Db 82 VTSBGDHYHYNGKVPYDAIIEELLMDPNYHFKGSDVINELDGYIVKNGVYVYVLK 141
QY 120 DAAHADNVRTKDEINRQKQEHVKD--NEKVNNSVA-----VARSQGRYTTNDQYV 167
Db 142 PGSKRKNIRTKQIAEQVAKGTEKEKGLAQVAHLSKEEVAANVEAKRQGRYTTDQGYI 201
QY 168 FNPADIETDNAYIVPHGCHYHYIPKSDLSASELAQAHAHLA---GKNMPSQLSYSGT 224
Db 202 FSPDTIIDDLGDAYLVPHGNHYHYIPKDLSPSELAQAAYWSQKGRGAPRS--DYRPT 259
QY 225 ASDNNTQSVAKGSTSKPAN-----KSENQLQSLLKEL 255
Db 260 PAPGRKAPIDVTPNPGQHQPDNGYHPAPPRPNDASQNKHQHDEKFGKTFKELLDQ 319
QY 256 YDPSAQRYSQSDGLVDFDPAKIIISRTPNQVAIPHGDHYHYIPKSDLSALEEKIARWPI 315
Db 320 HRLDLKYRHEEDGLIFETQVKSNAFYVPHGDHYHYIPKSDLSALEEKIARWPI 379
QY 316 GTGSTVSTNAKNEVSSLSGSSSPSSSLTTSKELS-----SASD 355
Db 380 QTDNDNS-----GSDHSKPS-----DKEVTHFLGHRKAYKGLDGKPYDTS 423
QY 356 GYIFNPKDIVEETATAYIVRHGDHYHYIPKSDLSALEEKIARWPI 413
Db 424 AYVFSKESHSVDKSGVTAKHGDFHYI--GFELEQYELDEVANVWVAKQGADELVAALD 482
QY 414 HEKHEEDGYGFD---ANRIIAEDESQFVNSHGDHNYHFKKDLTEEGTKAAQKHLEEVK 469
Db 483 QEQGKEKPL-FDTKKVSRKVTGKGVYIMPDKGDIYFARYQLDITQIAFAEQELMLK 541
QY 470 TSHNGLDLSLSSHEQDYPGNKAKEMKDLKLEEKIAGIMKQYGVKRESIVVNEKKNALY 529
Db 542 KKHRYVDI-----VDTGIEPLAVDVSSLPNAGNATYDTGSSFVI-P 583
QY 530 HGDHHRADP---IDERKPYGIGHSHSNYELRPE--EGVAKKEGKNVYTGELTNVNL 584
Db 584 HIDHIVVPSYSLTRNQIATI-----KYVMQHPEVPRPDVNSKPGHE--ESGSIPIPNVPLD 637
QY 585 KSTFNQNFLLANGQKRVSFSPPELEKGLINMLKLTDPDGKLVLEKVSQKVFGEVY 644
Db 638 KRAGMFWQI-----IHSABEVQKALAE--RFAAPDGYIFD----- 672
QY 645 NIANFELDQYLPGLQTFKTYIASKDYPEYSYDGTFTVTSLAYKNASQTIFFYFPHAGTY 704
Db 673 -----PROVLAKETP-----VWKGDSFISIPA----- 694
QY 705 LRVNPOFAVPGTDALVRVDFEFGNAYLNNYKVEIKLPIPKLNQGTTRTAGNKIPVT 764
Db 695 -----DGSSLRITNSDL- 707
QY 765 EMANAYLDNQSYIIVEVPTILEKEN-----QTDKPSILPQFKNKQAQBSKDLKVEEPRKT 819
Db 708 -----SOAEWQAQELLAKKNAGDATTDKPEEQQ--ADKSNEN-----QQPSE 750
QY 820 SEKVEKEKLSGTNSTSNTLEVEPTVD-----PVOEKYAKFAESYGMKLENVFNMDGTI 875
Db 751 ASKEEKE-----SDDFIDSLPDYGLDRATLEDHINQLAQKANIDPKYLIFQPEG-V 800
QY 876 ELYLPSGEVI 885

Db 801 QFYKNKGELV 810
RESULT 10
Q93GT5 PRELIMINARY; PRT; 825 AA.
AC Q93GT5
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Histidine triad protein of group A streptococci.
OS HTPA.
GN Streptococcus pyogenes
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-9;
RA Terao Y., Kawabata S., Hamada S.;
RT "Characterization of a novel histidine triad protein of group A
RT streptococci";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB073859; BAB71774.1;
SQ SEQUENCE 825 AA; 92623 MW; DE4ECC199181DFFB CRC64;
Query Match 16.9%; Score 897; DB 2; Length 825;
Best Local Similarity 26.5%; Pred. No. 2.1e-34;
Matches 258; Conservative 152; Mismatches 293; Indels 270; Gaps 31;
QY 2 AYALNQRSCQENKNNRVSYVDGSSQSKS--ENLTPDQVSQKEGIAQAEQIVIKITDQGY 59
Db 22 SYQLGKHGMSATKNDQIAYIDDSKGRAPKTNKMDQISAEGBISAEQIVVITDQGY 81
QY 60 VTSBGDHYHYNGKVPYDALFSEELLMDPNYQDKADIVNEKGGYIIKVDGKYVYVLK 119
Db 82 VTSBGDHYHYNGKVPYDAIIEELLMDPNYHFKGSDVINELDGYIVKNGVYVYVLK 141
QY 120 DAAHADNVRTKDEINRQKQEHVKD--NEKVNNSVA-----VARSQGRYTTNDQYV 167
Db 142 PGSKRKNIRTKQIAEQVAKGTEKEKGLAQVAHLSKEEVAANVEAKRQGRYTTDQGYI 201
QY 168 FNPADIETDNAYIVPHGCHYHYIPKSDLSASELAQAHAHLA---GKNMPSQLSYSGT 224
Db 202 FSPDTIIDDLGDAYLVPHGNHYHYIPKDLSPSELAQAAYWSQKGRGAPRSYRPTA 261
QY 225 ASDNNTQSVAKGSTSKPAN-----KSENQLQSLLKEL 255
Db 262 PAPGRKAPIDVTPNPGQHQPDNGYHPAPPRPNDASQNKHQHDEKFGKTFKELLDQ 319
QY 256 YDPSAQRYSQSDGLVDFDPAKIIISRTPNQVAIPHGDHYHYIPKSDLSALEEKIA-RWVPI 314
Db 322 HRLDLKYRHEEDGLIFETQVKSNAFYVPHGDHYHYIPKSDLSALEEKIA-RWVPI 379
QY 315 GTGSTVSTNAKNEVSSLSGSSSPSSSLTTSKELS-----SAS 354
Db 380 --AGTETDDSS-----GSDHSKPS-----DKEVTHFLGHRKAYKGLDGKPYDTS 424
QY 355 GYIFNPKDIVEETATAYIVRHGDHYHYIPKSDLSALEEKIARWPI 412
Db 425 DAYVFSKESHSVDKSGVTAKHGDFHYI--GFELEQYELDEVANVWVAKQGADELVAAL 483
QY 413 SHEKHEEDGYGFD---ANRIIAEDESQFVNSHGDHNYHFKKDLTEEGTKAAQKHLEEV 468
Db 484 DOEQGKEKPL-FDTKKVSRKVTGKGVYIMPDKGDIYFARYQLDITQIAFAEQELMLK 542
QY 469 KTSNGLDLSLSSHEQDYPGNKAKEMKDLKLEEKIAGIMKQYGVKRESIVVNEKKNALY 528
Db 543 DKKHRYVDI-----VDTGIEPLAVDVSSLPNAGNATYDTGSSFVI- 584
QY 529 PHGHHHRADPI-----DEHKPVGIGHSHSNYELRPE--EGVAKKEGKNVYTGELTNV 581

Db 585 PHIDHVVVPSWLTDRDIAI-----KYVMOHPEVRPDINSKPGHE-ESGSVIPNVT 636
Qy 582 NLLKNSFTNNQNFILANQKRVSPFFPELEKLGINMLVKLITPDGKVLKSVGKVFGE 641
Db 637 PLDKRAGPNWOI-----IHSAEVOKALAEG---RFAIPDGVIFD-----674
Qy 642 GVCNIANFELQOYLPFGOTFKYITASKDYPEVSDGTFTVPTSLAYKNASOTIIFYPFHAG 701
Db 675 -----PDVLAKETF-----VWKDGSFIPRA-----696
Qy 702 DTYLRVNPQFAVPKGTDALVRVDFEHGNAYLENNYKVGELKLPKLNQGTTRTAGNKI 761
Db 697 -----DGSRLTINKS 707
Qy 762 PVTMANAYLDNQSYIIVEPILEKEN-----QTDKPSILPQFKNKAQENSKLDERVEE 816
Db 708 DL-----SOAEWQAQALAKKNAGDATDIDKP-----KEQQAQDKSNENQOP 750
Qy 817 PTKSEKVEKELSETGNSNSTLEEVPTVD-----PVEKVAKAESYGMKLENVFNMD 872
Db 751 SEASKEEKE-----SDDFIDSLPDYGLDRATLEDHINQAQKANIDPKYILFOPe 801
Qy 873 GTIELVLPSEVI 885
Db 802 G-VOFYKNGELV 813
RESULT 11
Q9AE21 PRELIMINARY; PRT; 289 AA.
AC Q9AE21
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE Hypothetical 32.0 kDa protein (Fragment).
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5531;
RX MEDLINE=21172873; PubMed=11274116;
RA Granlund M., Michel F., Norgren M.;
RT "Mutually exclusive distribution of ts1548 and GBS11, an active group
II intron identified in human isolates of group b streptococci.";
RL J. Bacteriol. 183:2560-2569(2001).
DR EMBL; AJ290952; CAC35985.1; -
KW Hypothetical protein.
FT NON_TER 289
SQ SEQUENCE 289 AA; 32043 MW; A15A8588EA8140E4 CRC64;
Query Match 12.4%; Score 661.5; DB 2; Length 289;
Best Local Similarity 50.6%; Pred. No. 6a-24; Mismatches 19; Gaps 5;
Matches 130; Conservative 42; Indels 19; Gaps 5;
Qy 2 AYALNHRSOENKDNRVSYDVSQSQRK--ENLPTDQVSQKEGIAQVVIKVTDOQY 59
Db 22 SYQLKHHMGLATKQIAIYDSDGKVRAPKTKNTMDQISAEIGISAEQIVVKITDQY 81
Qy 60 VTSBGHDHYHNGKVPYDALFSELLMKDNYQLKADIVNEVKGYYIIVKDGKYYVYLK 119
Db 82 VTSBGHDHYHNGKVPYDAIISELLMTDNYRFDKSDVINEIDGVIKVNNGYVYVYLK 141
Qy 120 DAHADNVRTKDEINROKQEHVKD-NEKVNVA-----VARSGRYTTNDGVY 167
Db 142 PGSRRKIRIKQQLAEQVAKGTKEAKKGLAQVAHLSEKEVAVNEAKRGRYTTDDGYI 201
Qy 168 FNPADITEDTGNAYIVPHGGHYIPKPSDLSASELAHAAKLAH---GKNQOPSQLSYSST 224
Db 202 FSPDTIIDLDGDAYLVPHGNHYIPKPKDLSPELAQAQYWSQKQGRGAPRS--DYRPT 259
Qy 225 ASDNNTOSVAKGSTSKP 241

Db 260 PAPGRKRAPIDVTPNP 276
RESULT 12
Q99276 PRELIMINARY; PRT; 792 AA.
ID Q99276
AC Q99276;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE Putative intercalin A precursor.
GN INLA OR SPX1361.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21132684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Keaton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006574; AAK34188.1; -
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR Pfam; PF00560; LRR; 4.
DR SMART; SM00370; LRR; 4.
KW Complete proteome.
SQ SEQUENCE 792 AA; 87458 MW; 9D5E32288485ACE0 CRC64;
Query Match 5.5%; Score 292.5; DB 16; Length 792;
Best Local Similarity 21.4%; Pred. No. 6.2e-06;
Matches 185; Conservative 113; Mismatches 285; Indels 281; Gaps 45;
Qy 217 SOLSVSTASDNN-----TQSVAKGSTS---KPAKSENLOQLLKL--YDSPSAQRY 265
Db 16 SOLTILACQSRNGTYPIKTKQSRKGMTSNKIKPIKSKKTKNTHKGAVGVEPT-----70
Qy 266 ESDGLVF-DPAKILSRTPNGVAIPGHGHHYPIPYSKLSALEEKIARMVPISGTGVSTN 324
Db 71 -DDGFILTKDSKLSKTDQGLIVDHDGHSHEIFVADLKG--SPEYILIP---KGASL---121
Qy 325 AKPNEVSVSLGSLSNPSLTTSKELSSASDGYIFNPKDIVETATAYIVRHGHHFYIP 384
Db 122 AKP-----AVQRAASQGTSK-VADPHHHYEFNPADIVAEDALGYTVRHDDHFEYL 172
Qy 385 KSNQIGQDTLPNNSLAT--PSPSLPINPTSH-----EKHEEDGYGDANRIAEDESGF 437
Db 173 KSSLSGQTQAQAKQVATRLPQTSSLVSTATANGIPGLHFTSDGFGFNQGVGTVDKSI 232
Qy 438 VMSHGDHNYHFFKDLTE-----EQAKAAQKHLVEEVKTSNGLDLSLSSEHQDYPG 487
Db 233 LYVDHCHLHPIISFADLRGGWAHVADQYDPAKAEKPAETHQT-----275
Qy 488 NAKEMKDKLDKTEEKIAGIMQYGVYKRESI---VYNKEKNATIIYPHGDH-----533
Db 276 --PELSEREKEYQELAVLAELKLGIDPSTIKRVETQDQKLGLEYPHDHAHVLMLSIEI 333
Qy 534 -----HHADPIDHKKPVG-----IGHSHSNYELFKPEEGVAKKE 567
Db 334 GKIDPDPHAIEHARELEXHK-VGMDTLRALGDFDEVIDIVRTHDAPTFPSNE---KD 388
Qy 568 GNKVTGELTWNWNL---LKNSTFNQNFNFTLANGKRVSFSPPELE-----KLGII 617
Db 389 PNMM--KEWLATVIKLDLSGRKDPQLQRKGLSLPNLETGLIGFTPIKDISPVLQFKLKQ 446
Qy 618 NMLVKLITPDGKVLKSVGKVGEGVGNIANFELDQPLPGQTFKTYTATSKDYPVSYDG 677

```

Db 447 LLMTKGTVDYRFLD-----NNPQLEG-----IDISQNNLKDISELS 483
Qy 678 TFTPPTSLAYKASQTIFYPFHAGTYLR-VNPQFAVPKGTDAIVRVDFEFCNAYLENN 736
Db 484 KYKNUFLVA-----AANGIEDIRPLOGLP-NLKFLV-----LNN 518
Qy 737 YKVGEEKLPKLNQGTTRTAGNKIPVTFMANAYLDNSTYIVEV-PFILEKENQT-----790
Db 519 -KISDLS-PLASLHQ-----LQELHIDNQ--ITDLSPVSHRESLIVVDLS 560
Qy 791 -----DPFST-----LQFKKKAQENSKLDEKVEEPT 819
Db 561 RNADVLAFLQAPKLETLMVNDTKVSHLDLKNPNLSLSINRAQLQSL--EGTEASSV 618
Qy 820 SEKVEKEKLSGTNSTSTLEVEPTVDVQEVKAKFAESYGMKLENVFNMDG-----873
Db 619 IVRVAE-----GNQIKSLVLD-----KQSLTFLDVTG-----NQLTSLEGVNNETA 662
Qy 874 -----TIELYLSGEV-----TKN-----MADF--TGEAPQNGENKPS--ENG 909
Db 663 LDILSVSKNQLTNVNLSEKNTVITIDISHNNISLADKLNQHIPEAIKKNFPAVEGS 722
Qy 910 KYSTGTVENOPT-ENKPADSLPEA 932
Db 723 MYGNGTAEEKAMATKAKESAQEA 746

RESULT 13
Q97QP7
ID Q97QP7 PRELIMINARY; PRT; 2004 AA.
AC Q97QP7
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Immunoglobulin A1 protease.
GN Sp1154.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetelin H.; Nelson K.E.; Paulsen I.T.; Eisen J.A.; Read T.D.;
RA Peterson S.; Heidelberg J.; Deboy R.T.; Haft D.H.; Dodson R.J.;
RA Durkin A.S.; Gwinn M.; Kolonay J.F.; Nelson W.C.; Peterson J.D.;
RA Umayam L.A.; White O.; Salzberg S.L.; Lewis M.R.; Radune D.;
RA Holtzapple E.; Khouli H.; Wolf A.M.; Utterback T.R.; Hansen C.L.;
RA McDonald L.A.; Feldblyum T.V.; Angiuoli S.; Dickinson T.; Hickey E.K.;
RA Holt J.E.; Loftus B.J.; Yang F.; Smith H.O.; Venter J.C.;
RA Dougherty B.A.; Morrison D.A.; Hollingshead S.K.; Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AE007416; AAK75263.1;
DR MEROPS; M26.001;
DR TIGR; SP1154;
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001300; Zn_MiPeptdse.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMS; TIGR011167; LPXIG_anchor; 1.
DR TIGRFAMS; TIGR011168; YSIRK_signal; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease; Complete proteome.
SQ SEQUENCE 2004 AA; 223908 MW; 556BC6A1028D60A4 CRC64;

Query Match 4.5%; Score 238.5; DB 16; Length 2004;
Best Local Similarity 22.6%; Pred. No. 0.0081;
Matches 206; Conservative 110; Mismatches 284; Indels 313; Gaps 52;
Qy 186 GGHYHYPKSDLSASELAANKAHLAGKNQPSQLSY---SSTASDNNTQSVAKGSTSKPA 242

```

```

Db 47 GVHKYVADSELSSEE-----KKQLYDIPTYVNDDETYLV-----84
Qy 243 KNSNLOSLKELXDSPSAQRYSESDGLVFPDAKIISTPNGVAIPHGDHYHFIPYSKLS 302
Db 85 -YKLNQNLAEI--PNTGSKNERQALVAG-----ASIAAMGILT-----FAVSKKK 128
Qy 303 ALEKIAKRVPISTGS--TYSTNAKPNEVSVSLGSSNPSSLTTSKELSSASDGVIFN 360
Db 129 VNKTVLHLVIVAGLNGVLVSHALENHL-----LYN-----TYELTSGEK--LPL 176
Qy 361 PKDIVEETATAYIVRHGDHFYIPKSNQIGOPTLPNSLATPSPSLPINTPGTSHKEHEED 420
Db 177 PKETSGVTYIGYI-KEGKTTSESEVSQ-----KSSVATPT-----KQKV 216
Qy 421 GYGEDANRIIAEDSGFVMSHGDHNYHFFKKDLIEBQIAKAAKHLEEVKTSNHLDSLSS 480
Db 129 VNKTVLHLVIVAGLNGVLVSHALENHL-----LYN-----TYELTSGEK--LPL 176
Qy 361 PKDIVEETATAYIVRHGDHFYIPKSNQIGOPTLPNSLATPSPSLPINTPGTSHKEHEED 420
Db 177 PKETSGVTYIGYI-KEGKTTSESEVSQ-----KSSVATPT-----KQKV 216
Qy 421 GYGEDANRIIAEDSGFVMSHGDHNYHFFKKDLIEBQIAKAAKHLEEVKTSNHLDSLSS 480
Db 217 DYNTPNFV-----DHPSTVOAIQEQTPVSSTKP--TEVOVVEKPFSTELINPKRKEKQSDS 272
Qy 481 HEQYPGNAKEMKDLKKIEEKIAGINKQYGVKRESIVNKEKNAIYVPHGDHHAAPID 540
Db 273 QEQ-----LAHKNLETKEEKIS-----PKEKTGV-----NTLNPQD 305
Qy 541 E-----HKPVGIGHSHSNYELFPBEGVAKKEGNKYVTGEELTNVNLKNSTFNQNF 594
Db 306 EVLSGQLNKP-----ELLYREETME---TKIDFQEEI-----QENP 338
Qy 595 TLANGQKRVSFSPPELEKLGINN-LVKLTJ-----PDGKYLEKVS 636
Db 339 DLAEGTVRV-----KQEGKLGKVKVIRFISVNKEEVSREIVSTSTAPSRIVEKGT 392
Qy 637 K--VFGE-----GV-----GNIANFELDPQLPCQTFKYTIASKDYPEVSDGTFVTPT 683
Db 393 KTVIKQEPETGVEHKVQSGAIVEPAI-QPELP-----EAVVSDKGEPEVQP-----TLPE 443
Qy 684 SLAYKMASQTIYFPFHAGDTYLRVNPQ-----FAYPKGTDALRVDFDEHGNAYLENKY 739
Db 444 AVV-----TDKGET--EVQSPESPDVVSXKGEPEQVAPLPEYKGN-----I 482
Qy 740 GEIK--LPIPKL-NQGTTRTAGNKIPVTFMANAYLDNQSTYIVFPILEKENQDKPSIL 796
Db 483 BOVKPETPEVKTQGGPEKT--EEVPV-----KPEETPVNPNEGTTGTSI- 527
Qy 797 PQFRN---KAQENSKLDEKVEEPEKTSEK-----VEKELSETGNST 835
Db 528 -QEAENPVQPAEESTTNSEKV-SPDTSKNTGEVSSNPSSDTSVSGESNKPENHDSKNEN 585
Qy 836 SNSTLEEPTVDVQEVKAKFAESYGMKLENVFNMDGTIELYLPSSGEV IKKNADFTGE 895
Db 586 SEKTVEEVP-VNP-----NEGTVE-----GTSNQETE 611
Qy 896 APQNGENKPSSENGKV---STGTVENOPTENKP--ADSLPEAPNEKPKVPKPS-----TDN 946
Db 612 KPVQFAEETQTNNGKIANENTGETSVSNKPSDSKPPVEESNQPEKNGATKPKNSGNTTSEN 671
Qy 947 GMLNPE---GNVGDPMLDPALEAPAVDPVQEKLEKFTASYGLDVSIFNMDGTIELR 1003
Db 672 QCTPEPSNGNSTEDVSTESNTSNGNEEKQENE-----LDPKKVVEPEKTELELR 724
Qy 1004 LPSGEVTKKNLSD 1016
Db 725 -----NVSD 728

RESULT 14
Q963TL
ID Q963TL PRELIMINARY; PRT; 1078 AA.
AC Q963TL;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Glutamate-rich protein (Fragment).
GN GLURP.

```

```
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21313688; PubMed=11420113;
RX Theisen M., Thomas A.W., Jepsen S.;
RT "Cloning, nucleotide sequencing and analysis of the gene encoding the
RT glutamate-rich protein (GLURP) from Plasmodium reichenowi.";
RL Mol. Biochem. Parasitol. 115:269-273(2001).
DR EMEL; AF356828; AK40236.1;
FT NON_TER 1 1078
ET NON_TER 1078 1078
SQ SEQUENCE 1078 AA; 123581 MW; 8D388D88B223913C CRC64;

Query Match 4.4%; Score 235.5; DB 5; Length 1078;
Best Local Similarity 19.6%; Pred. No. 0.0047;
Matches 209; Conservative 157; Mismatches 380; Indels 321; Gaps 51;

QY 6 NOHRSQENKNNRVSYVDGSSQSKSENLTDPQVSQKEGIAEQIVIKITDQGYVTSBGD 65
DB 200 HEBKLDNHYEKNVFNHSSVNSQENLKPESFDEHLNISK-----KLEDDLDLHEHGN 255
QY 66 HYHYNGKVPYDALSFEEL-LMKDNYOL-----KDADIVNEVKGYYLIK 109
DB 256 SEHLKDEKIGNEFLYHENSIFNDIEQILNLPQEQTNVQEQLYNEKQNVKPKNSEIFS 315
QY 110 VPGKYYVYLKDAADNATKDEINRQKQ-----BHVK-----DNEKVN---148
DB 316 LD-----LKETTNDILNPQPLENIKQSESEINHVDHALPENIIDKLDNEKEHIDE 369
QY 149 --SNVAVASQGRYTTNDGYV-----PNPADI-----IEDTGNAYIVPHGG 187
DB 370 SQHNVNVLQEN---NINNQLPEQKANTIESEPNIDSEIILPENVEKEIIVDVPSPK 426
QY 188 HY-HYIPKSDLSASELAALAAKHAHAGKNNOPQSLSYSTASDNNNTOSVAKGSTKYPANKSE 246
DB 427 HFNHETLEQETSESEHEEA---VSEKNAHETVEHEEAVSQESNPE-----KADNDGNVSQ 478
QY 247 NLOSLLKELYDSPAQSYSESGLVDFPAKIIISRTPNGVAIPHGDHYHFIPYKSLSALEE 306
DB 479 NSNNELNE-----NEFVESKSEHEPAE-----NEESSLEE 509
QY 307 KIARMPISGTGTSTVETNAKPNNEVSSLSGSLNSPSSLTTSKELSDSGYTFNPKDIVE 366
DB 510 -----AHQBEIVPEQNNQESGESKLVNDDE-----GGF-----E 538
QY 367 ETATATVIRGHDFHYIPKSN--QIQOPLPNNSLATPSPSLPINPGTSHKEHEED--- 420
DB 539 EAHQEEIV-----PEQNNQESGESKLVN-----DEGGFEAAHQEEIVPE 578
QY 421 --GYGFDANRIIAEDSGFVMSHGD-----HNHYFFKKDIT----- 454
DB 579 QNNQESGESKLVNDDEGFEAAHNFSEESNSLHEHEEVESDESDEPPEHEEVSE 638
QY 455 EEOIKAAQKHLEEVKTSNGLDLSLSHBOYPGNAKEMKDLKKIETKTAGI-----MKQY 510
DB 639 ESNPEASENSESSIEEAHQ--EEIEQNDSELNELVESEKSEPAEHVEIVSEKSV 696
QY 511 GVKRESIVNKKENAIYIPGHDDHHDADPIDEH---KPVGTGCHSHSYELFKPEEGVAKKE 567
DB 697 SEPAEHVEIVSEKSV---SEPAEHVESVSEASPSSEHVESVSQSNNPESE---KKD 749
QY 568 G---NKYVTGEEITNVNLLKNTFNQNT---LANGQRVSESF-----PPELEKGLGI 617
DB 750 GPVPSKPFEEIEKVDQPKIVDLQIIIEPNFVLDQNPQEPFVPSFKYIEKVPSEENKHC- 808
QY 618 NMLVKLITPGKVKYKSGKVGFGVGGVGNATNFELDQPLPGQTFKVTIASKD-----YEE 672
DB 809 -----SVDPEVEEK-----ENVSEV-----VEKQNPQELFELPLKDKIEKIEPE 850
QY 673 VSYDGTFTVPTSLAYKNASOTIYFPFHAGDTYLRVNPQFAVPGTADALVRVDFEFGNAY 732
```

Search completed: May 13, 2003, 13:56:33
Job time : 55.2509 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:51:52 ; Search time 15.7951 Seconds
(without alignments)
1898.176 Million cell updates/sec

Title: US-09-471-255-55

Perfect score: 5315

Sequence: 1 CAYALNHRSGENKNNRVS.....TEFLPSEGVTKNLSDFIA 1019

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2338	44.0	447	4	US-08-961-083-182
2	2247	23.5	796	4	US-08-961-083-56
3	1228.5	23.1	763	4	US-08-961-083-66
4	257	4.8	10182	4	US-09-134-001C-3159
5	223	4.2	1964	2	US-08-790-912-3
6	223	4.2	2052	2	US-08-790-912-2
7	205.5	3.9	1183	2	US-08-447-031A-2
8	202	3.8	1848	4	US-08-236-791-6
9	202	3.8	1848	5	PCT-US93-10661A-6
10	200	3.8	571	4	US-08-961-083-4
11	198	3.7	851	4	US-09-071-035-326
12	198	3.7	851	4	US-09-071-035-330
13	198	3.7	851	4	US-09-071-035-334
14	186.5	3.5	3696	4	US-09-134-001C-5080
15	186	3.5	1231	3	US-08-904-263A-4
16	185.5	3.5	1507	6	5268270-2
17	178.5	3.4	2465	2	US-08-596-291-3
18	178.5	3.4	2465	3	US-09-100-804-3
19	177	3.3	1702	4	US-08-296-791-5
20	177	3.3	1702	5	PCT-US95-10661A-5
21	175.5	3.3	2466	3	US-09-080-855-12
22	175.5	3.3	2466	5	PCT-US94-03943-2
23	174.5	3.3	2485	4	US-09-290-640-46
24	173.5	3.3	1177	4	US-09-134-001C-5106
25	173	3.3	1780	1	US-08-769-309A-5
26	173	3.3	1780	3	US-08-994-570-5
27	171	3.2	984	1	US-08-242-932-2
28	171	3.2	984	5	PCT-US95-06111-2
29	171	3.2	1002	4	US-09-268-347-24
30	170.5	3.2	1043	3	US-08-928-361B-30
31	170.5	3.2	1721	3	US-08-700-651-5
32	170.5	3.2	1721	3	US-08-928-361B-6
33	169.5	3.2	1167	2	US-08-589-756-2
34	169.5	3.2	1167	4	US-09-206-800-2
35	169.5	3.2	1167	4	US-09-206-898-2
36	169.5	3.2	1167	4	US-08-568-459A-4
37	168.5	3.2	1435	2	US-08-487-826B-4
38	168.5	3.2	1435	2	US-09-210-268-4
39	168.5	3.2	1435	4	US-08-923-992A-4
40	167.5	3.2	1104	4	US-08-923-992A-8
41	166.5	3.1	1098	4	US-08-923-992A-6
42	166.5	3.1	1128	4	US-08-923-992A-10
43	166.5	3.1	1164	4	US-08-923-992A-2
44	164.5	3.1	1184	4	US-08-923-992A-2
45	164	3.1	962	4	US-09-071-035-246

ALIGNMENTS

RESULT 1

US-08-961-083-182

; Sequence 182, Application US/08961083

; Patent No. 6159469

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,083

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 182:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 447 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-961-083-182

Query Match 44.0%; Score 2338; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 6,7e-160;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LNHRSEKNNRVSVDGSSQSKSENLTPOVSKQEGIAEQIVIKITDGYVTSHG 64
Db 1 LNHRSEKNNRVSVDGSSQSKSENLTPOVSKQEGIAEQIVIKITDGYVTSHG 60

Query Match 23.5%; Score 1247; DB 4; Length 796;
Best Local Similarity 41.2%; Pred. No. 3.4e-81;
Matches 298; Conservative 95; Mismatches 180; Indels 150; Gaps 19;

QY 65 DHYHYNGKVPYDALFSEBLLMKDPNYOLKADIVNEVGKGYIIKVDGKYVYVYLKDAHA 124
DB 61 DHYHYNGKVPYDALFSEBLLMKDPNYOLKADIVNEVGKGYIIKVDGKYVYVYLKDAHA 120
QY 125 DNVRYKDEINROKQEHVDNEKXVNSNVAARSQGRYTTNDGVFNPAIIEEDTGNAYIVP 184
DB 121 DNVRYKDEINROKQEHVDNEKXVNSNVAARSQGRYTTNDGVFNPAIIEEDTGNAYIVP 180
QY 185 HGCHYHYIPKSDLSASELAARAHLAGNMOPSQLSYSTASDNNQTSVAKGSKPANK 244
DB 181 HGCHYHYIPKSDLSASELAARAHLAGNMOPSQLSYSTASDNNQTSVAKGSKPANK 240
QY 245 SENLOSLKELYDSPAORYSESDGLVFPDPAKIIISRTNGVAIPHGDHYHFIPTSKLSAL 304
DB 241 SENLOSLKELYDSPAORYSESDGLVFPDPAKIIISRTNGVAIPHGDHYHFIPTSKLSAL 300
QY 305 EKIARMPISGTGSTVSTNAKNEPNEVSSLSLSPSLTTSKELSSASDGYIFNPKDI 364
DB 301 EKIARMPISGTGSTVSTNAKNEPNEVSSLSLSPSLTTSKELSSASDGYIFNPKDI 360
QY 365 VETATAYIVRGDHPHYIPKSNQIQPTLPNNSLATPSLSIPINPGTSHKHEEDGYGF 424
DB 361 VETATAYIVRGDHPHYIPKSNQIQPTLPNNSLATPSLSIPINPGTSHKHEEDGYGF 420
QY 425 DANRIIAEDESFGVMGSHGDNHYFFKK 451
DB 421 DANRIIAEDESFGVMGSHGDNHYFFKK 447

RESULT 2

US-08-961-083-56

; Sequence 56, Application US/08961083

; Patent No. 6159469

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,083

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 56:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 796 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-961-083-56

Query Match 23.5%; Score 1247; DB 4; Length 796;
Best Local Similarity 41.2%; Pred. No. 3.4e-81;
Matches 298; Conservative 95; Mismatches 180; Indels 150; Gaps 19;

QY 2 AYALNHRSEKNDNRVSYVDGSSQSKSENLTPOVQSKQEGIAEQIVIKITDGYVT 61
DB 1 SYELGLYQARTVAKENRVSVIDGQATKTEINTPDVSKREGIAEQIVIKITDGYVT 60
QY 62 SHGDHYHYNGKVPYDALFSEBLLMKDPNYOLKADIVNEVGKGYIIKVDGKYVYVYLKDA 121
DB 61 SHGDHYHYNGKVPYDALFSEBLLMKDPNYOLKADIVNEVGKGYIIKVDGKYVYVYLKDA 120
QY 122 AHADNVRTKEINRQKQEHVKDNE---KVNNSNVAARSQGRYTTNDGVFNPAIIEEDT 177
DB 121 AHADNVRTKEINRQKQEHVREGGTTPRDGAVALARSGQRYTTDDGYIFNASDIIEEDT 180
QY 178 GNAYIVPHGCHYHYIPKSDLSASELAARAHLAGNMOPSQLSYSTASDN-----228
DB 181 GDAYIVPHGCHYHYIPKSDLSASELAARAHLAGNMOPSQLSYSTASDN-----240
QY 229 -----NTOSVAKGSTSKPANKSENLOSLKELYDSPAORYSESDGLVFPDPAKIIISRT 281
DB 241 VSNPGTTNTNTSNNSTNSQASQSDNDIDSLKQYKLPKLSQRHVESDGLVFPDPAQITSRT 300
QY 282 PNGVAIPHGDHYHFIPTSKLSALEEKIARMP-----313
DB 301 ARGVAVPHGDHYHFIPTSKLSALEEKIARMP-----360
QY 314 -----ISGTGSTVSTNAK-----PNEVSSLSLSPSLTTSKELSSASDGYIFNPKDI 420
DB 361 POPAPNLKIDSSNSLSQVLRKYCEGVYFEEKGISRVYFAKDLPSSETVKNLESKLSKQES 420
QY 344 -----LTTSKELSSASDGYIFNPK--DIVETATAYIVPHG--DHFHYIPK-----SNQIG 390
DB 421 VSHTLTAKENVAPRDQEFYDKAYNLLTEAHKALFXNKGNSDFQALDKLLERLNDESTN 480
QY 391 OPTLPNNSLATPSP-SLPINPGTSHKHE-----EDGYGFDAKRIIAE 432
DB 481 KEKLVDDLLAFAPITHPERLKGKNSQIEYDEVRITLAQLADKYTTSDGYIFDEHDIISD 540
QY 433 DESGFVMSHGDNHYFFKKDLTERQIAAKAKHLEE-----VKTSHNGLDLSLSSH 481
DB 541 EGDAYVTPHMGHSHWIGKDSLSLDEKVAQAAYTKKEGILPSPDADYKANPTG-DSAAA 599
QY 482 EODYPGNAKEMKDLKTEEKIAGIMQYGVKRESIVVNKEKNAILYPHGDHHDADP---538
DB 600 YNRVKG-----EKRIPLVRLPYV--EHTVEVKNGLII-PHKDHYHNKFAW 644
QY 539 IDEH---KPVGIG---HSHSNYELFKPEEGVAKKEKNKYVTGELTNVNLKNSTFNN 591
DB 645 FDDHTYKAPNGYTTLEDLFATIKYVVEHPDERPHSNDG----WGNASEHVLGKDKHSDPN 700
QY 592 QNF 594
DB 701 KNF 703

RESULT 3

US-08-961-083-66

; Sequence 66, Application US/08961083

; Patent No. 6159469

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

; NUMBER OF SEQUENCES: 452

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:


```

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/333
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961.083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: B9340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 763 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-66

```

Query Match 23.1%; Score 1228.5; DB 4; Length 763;
Best Local Similarity 40.5%; Pred. No. 6.8e-80;
Matches 296; Conservative 101; Mismatches 185; Indels 149;

QY		1	CAYALNOHRS-QENKONNRVSYVDGQSOSOKSENLPDQVSQKEGIQAEOIVIKITDOGY	59
Db	: :	1	CSVELGRHQAGOVKKESNRYSIDGOAOGAKBENTLDPDESKREGINAEOXVKIITDOGY	60
QY		60	VTSHGDRHYHYNGKVPYDALFSEELMKDPNYOLKDADIVNEKVGYILKDVGGYYVYLK	119
Db	: :	61	VTSHGDRHYHYNGKVPYDAISELLMKDPNYOLKESDIVNETKGGVIVKYNGKYVVYLK	120
QY		120	DAHADNVRTKDINROKQBHVXD-NEKVNSNVAVARSOGRIYTNDGYVFNPADIIEDTG	178
Db	: :	121	DAHADNVIRTKEEIKRQKBQRSHNHRSRADNAVAAARAGRYTTDDGYIFNASDIIEDTG	180
QY		179	NAYIVPHGGHYHIYPKSDJLSASELAALKAHLAGKNWPOLSYSSTASDNNTQ--SVAK	235
Db	: :	181	-DAYIVPHGDHYHIYPNELSASELAAREAYWNGK--QGSRPSSSSSYNNANPAQLUSENH	238
QY		236	GSTSKEPA---NKSENLOSLKELYDSPSAQRYSESGLVFPDAKIISRTPNGVAIPGHQH	292
Db	: :	239	NLTVTPTYHQNCENTSULLRELYAPLEPSHERVESGLIFDPAQITSRTARGVAVPHGNH	298
QY		293	YHTPYKSLSALEBKIAARMVIPSGTGSTVNAPKN-----EYVSSLGSLS	338
Db	: :	299	YHTPYEQNSELEKRARIPIPLAYSRNHWVPDSRPQPSPQTPEPSPDQAPNPQAP	358
QY		339	SNFSSUTTSVE-LUSSADGYIFN-----PKDIVEETAT---AYIVRHGDHFHY---	382
Db	: :	359	SNPIDEKLVKAERYKVGDGYPFGENVGSRYIYPAKDLSEAETAACIDKLAKESLSHKGA	418
QY		383	----IPKSN-----QIGQPTLPNNSLATSPSL-----	406
Db	: :	419	KKTDLSSOREFYKAYDLLARHQDLLONGQOVDFFALDNLLERLKDVXSCKVLVXD	478
QY		407	----PI-----NPGTSEKHIE-----EDGYGFANDRIIAEDESFGYM	439
Db	: :	479	IILAFLAPIRHPERBLRGPNQAITYTDDBEIQVAKLAGKYTTDEDGYIFPPROIITSDGDYVT	538
QY		440	SHGDHNHYFFKBDLTEOIKAACKHELEVKTSNHGJLDSLSSHEDQYPGNAKEWKOLDK--	497
Db	: :	539	PMTWHSHWTJKDLSAEARAADAQAYAKE-----KGLTPPSTOHQD-SGMT-SAKGAETY	591
QY		498	--KIEEKTAGIMQYGVARRESTVWNKENAIYIPHGDHHHADPIDEHKPVGIGHSHSY	554

```

Db 592 NRVAAKKVLDRPYNLQ-----YTVEKNGSLIIPHYDHYH-----NIKF 633
Qy 555 ELFPDEGVAKEGNKVKYTGELTNVNVLL-----KNSTNNONFTL-----ANGQR 602
Db 634 EWF--DEGL--YEAPKGVTLLELATVKYVVEHNERPHSDNGFGNASHVQRNKNQAD 689
Qy 603 VSFSEPPPELEK 613
Db 690 TNQTEKPFSEK 700

RESULT 4
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: staphylococcus epidermidis
; US-09-134-001C-3159

```

Query Match 4.8%; Score 257; DB 4; Length 10182;
Best Local Similarity 19.3%; Pred. NO. 1.9e-08;
Matches 226; Conservative 183; Mismatches 455; Indels 308; Gaps 52;

Qy	5	LNHRSOENKDNRRVSYVDGSSOSSOKSEMLTPDQVSKQEGI-QABEQVIKITDQGYTSH	63
Db	5218	INEDANQOQAYSAIASAALKN--KSONPELDKVTIQAINNINSAIINNLNGEAKLT--	5273
Qy	64	GDHYHYNGVPVYDALFSEELLMKDPNYOLKDADIVNEVKGQYIHKVDGKYVYVKDAAH	123
Db	5274	-----KAKEDAVASINN-----SGLTNQK-----TKENQA	5300
Qy	124	ADNVRPKDEI-NRQKEHVKD-----NEKVMNSVAVARSQGRYTTNDGYVFNPAIDIED	170
Db	5301	VNGAOTERDQVANKLRDAEALDSMOTLRDLVNNQNAI-HSTSNFNEIDSTQKNTYDRAID	5359
Qy	177	TGNAYTVPHGGHYHTPKS--DLSASELAAAKAHLAG-----KNMOPSOLESYST	224
Db	5360	NGSTYIT--GOHPELNKSTIDQITISRTINKNDLHVKEQLQRDKGTANQIOLGVLND	5417
Qy	225	ASDNNTQSVAKGSTKRP-----ANKSENLOSLKEL-----YDSESAQRY	264
Db	5418	PQSGEESLVNGSNTSEVEEHLNEAKSLLNNAMKLRDKVAEKTNVKOSSDYINDSTEHQ	5477
Qy	265	SESDGVLVFPQAKIISRTPTGVAIPHGDHYHTFYPKLSALBEKIAKMPISGT--GTVS	322
Db	5478	RGYDQALQEAENIIEIGN-----PTLNKSEIQKICQLTDAQNALQGSHELL	5524
Qy	323	TNAKPNVYSSLOSLSNPFSSITTSKEI:SSASDGYIFNPKDQIVETA:YAYIVRHGDHFHY	382
Db	5525	EEAKNN-----ATTGINKLTALNDQAORAIENVAQQT-----	5558
Qy	383	IPKSNIGTOPTLPNNSLATPSP:SLNP:GTSI:HEKHEEDGY-----GFDANRI	429
Db	5559	IPAVNQ--QULT--DREINTAMQALRDKVGOQNNVHQOSNYFNEDEQPKHNVDSNVQAGQT	5615
Qy	430	IADESGFVMSGHDHNNHYFFKKDLTEQIKAQK-HLEVKTSH--NGLDLSL-----SHE	482
Db	5616	IIDKLQCPILMNKEIDQAINQNTQTALSGENKLUHQDQESTNRQIEGSLSLTAQNAE	5675

QY 483 QDYPGNAKEMKDLKKI-----EKKTA-----GIMKQGVKRESIVNKKNAIIPHG 532
Db 5676 KDLVNOAKTRIDVAOKLAAAEINISAMENLDRGQNKEDIRSSGAYINADTKVTAYDQA 5735
QY 533 HHADPIDEHKVIGISHSNYELFKPEEGYAKKEGKRVYTGEEELTNVYNLLKNS----- 587
Db 5736 LQNAENIINATP-----NVELNK-----AIEQALS RVQQAQDLGVOQLANAKQOAT 5784
QY 588 -TFNQNFTLANGKRVSEFPPELEKLGINMLVKLITPCKVLEKVS-----KV 638
Db 5785 QVINGLIN-SLNDGQR-----ELNLLINSANTRTKVOEELNKATELNHAMEA 5830
QY 639 FGEGVGNIANFELDQPLPGQTFKYTIASKDPEV-SYDGTFTVPTSLAYKMASQTIETP 697
Db 5831 LNSVONVDQVQSSNY-----NEDQPEOHNDN-----AVNEAQATI--- 5869
QY 698 FRAGDTYLRVNPQFVAPKGTDALRVFDEFHGNAYLENNYKVGEIKL----- 744
Db 5870 ---NNAQPVLDKLAIERLTQVNTTKDALHGAQKLTQDQAAETGIRGLTSLNBPQKNA 5926
QY 745 PIPKLNQGTTRAGNKI---PVTFMANAYL-----DNQSTYIVE----- 780
Db 5927 EVAKVTAATREDEVRNIREATLTOTLMLGLRKSINKDNIDKNSKYINEDHDQQAAYDN 5986
QY 781 -----VPILEKENQTDKPSILPQFRKNAQENSKL--DEKVEEPTSEK---VEKEKLS 830
Db 5987 AVNAQOVIDEQTALSSDTINOLANAVTOAKSNLHGDTKLQHKDKSAKQTTIAQLQNLNS 6046
QY 831 TGNSTNSLIEEPTVDPVQEVKFAESYGMKLENVLFNMDGTIELY---LPSGEVI-- 885
Db 6047 AQKHMEDSLIDNESTRTOVQHOLTE-----AQALDGLMGALKESIKDYNIVNSGNYINA 6101
QY 886 ---KKNMADFTGAPOG---NGENKPSN--GKYSTG--TVENOPTENKPADSLPEAPN--E 935
Db 6102 EPSKKAYDAVQNAQNLINGNPTINKGNVTTATQTVKNTKDALDGDHRLKEAKNNAN 6161
QY 936 KPYKPNSTNDGMLNPEGNGVSDPMDLPALEAPVDPVQEKLEKFTASYGLGLDSVIFN 995
Db 6162 QTRNLNLSNNAQDAEKNL-----VNSASTLEQVQOQLQ---TAQO---LQNAME 6207
QY 996 MDGTIELR-----LPSGEVIRKKNLSDFI 1018
Db 6208 LRSIAKKDQVKADSKYLNEDPOIKQNYDDAV 6239

RESULT 5
US-08-790-912-3
Sequence 3, Application US/08790912
Patent No. 5976542
GENERAL INFORMATION:
APPLICANT: Weiser, Jeffrey N.
APPLICANT: Plaut, Andrew G.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: 1601 Market Street, 36th Floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/790,912
FILING DATE: 29-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,939

FILING DATE: 23-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Leary, Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-4U1
TELEPHONE: (215) 567-2020
TELEFAX: (215) 567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1964 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-790-912-3

Query Match 4.2%; Score 223; DB 2; Length 1964;
Best Local Similarity 22.8%; Pred. NO. 4.3e-07;
Matches 191; Conservative 98; Mismatches 264; Indels 292; Gaps 48;

QY 186 GGHVHYIPKSDLSASELAARAKHLAKGNQPSQLSY---SSTASDNNQTSVAKGSTSKPA 242
Db 47 GVHYKYVADSELSSEE-----KKQLVYDIPYVENDDETYLV----- 84
QY 243 NKSENLQSLKELYDPSAORYSESDGLVDPKAIISRTENGVAIPHGDHYHIFYPYKLS 302
Db 85 -YKLNQNLAEEL---PNTGSKNERQALVAG---ASLAALGILI-----FAVSKKK 128
QY 303 ALEEKIARMPVISCIGS--TVSTNAKPNVSVSSLSGSSNPSSLLTSKELSSASDGYIFN 360
Db 129 VKNKTVLHLVUAGNGVLSVVALENHLL-----LNYN-----TDYELTSEK--LPL 176
QY 361 PKDIVEETATAYTVRHGDHFIYIPKSNQIOGPTLPNNSLATPSPSLPINFGTSHKHEED 420
Db 177 PKEISGVYIGVI--KEGKTTSDSEVSNO-----EKSAATPT-----KQKV 216
QY 421 GYGEDANPIIAEDSEGFVNSHGDHNYFFKKDLTEQIKAAQKHLBEVKTSHNGLDLSLS 480
Db 217 DYNVTPNEV---DHPSTVOAIOEOTPVSSIKP--TEVQVVEKPFSTELINPKKEKQSSDS 272
QY 481 HEQDYPGNAKEMKDLKKIEEKIAGIMKQGVKRESIVNKKNAIIPHGDDHHADPID 540
Db 273 QEQ-----LAEHKMLETKKEEKIS-----PKETGV-----NTLNPD 305
QY 541 E-----HKPVGIGHSHSNYELFKPEEGYAKKEGKRVYTGEEELTNVYNLLKNSFTNNQNF 594
Db 306 EYLSQQLNKP-----ELLYREETIE---TKIDFOEEI-----QENP 338
QY 595 TLANGKRVSEFPPELEKLGINN--LVKLIT-----PDGKYLEKVS 636
Db 339 DLAEQTVRV-----KQSEKLGKKEVIRFESVNKEEVSREIVSTSTTAPSPRIVEKGT 392
QY 637 K--VFGE-----GV-----GNIANFELDQPLPGQTFKYTIASKDPEVSYDGTFTVPT 683
Db 393 KTVIKKEQPTGVHEKQVQSGAIVEPAI--QPELP-----EAVVSORGEPEVOP---TLEPE 443
QY 684 SLAYKMASQTIETPFPFHAGDTYLRVNPQ-----FAVPKGTDALRVFDEFHGNAYLENNYKV 739
Db 444 AV-----TDKGET---EVQPESPDVTVDKGEQVAPLPEYKGN-----I 482
QY 740 GEIK--LPIPKL--NOGTTTACNKTPTVTFMANAYLDNOSTYIVFVPILEKENQTDKPSIL 796
Db 483 EGVKPTPVKTKEQGPEKT---EVPV-----KPTETPVNPNNEGTEGTSI-- 527
QY 797 PQFKN---KAQENSKLDEKVEEPTSEKV-----EKEKLSSETGNS 836
Db 528 -QEAENPVQPAEESTINSEKVSPTSSSENTGEVSNPSSDTSYVSGESKNKPEHNSKNNS 586
QY 837 NSTLEEVPTVDPVQEVKFAESYGMKLENVLFNMDGTIELYLPSEGVIKKNMADFTGEA 896
Db 587 EKTVEEVP-VND-----NEGTV-----GTSNOETEK 612

```

QY      897  POGNGENKPSKGVK---STGTVENQPTENKP--ADSLPEARNEKEPKVPENS-----TDNG 944
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      613  PVOPAEETOTNSKIANENTGEVSNKPSDSKPEVSESNQPEAKNGIATKPKNSGNTTSENG 672
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
QY      948  MLNPE 952
      |   |
Db      673  QTEPE 677

RESULT 6
US-08-790-912-2
; Sequence 2, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790.912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-4U1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-790-912-2

Query Match 4.2%; Score 223; DB 2; Length 2052;
Best Local Similarity 22.6%; Pred No. 4,6e-07;
Matches 191; Conservative 98; Mismatches 264; Indels 292; Gaps

QY      186  GGHYHYIPKSDLSASELAAKAHLAGKNQPSQLSY---SSTASDNTQSVAKGSTRKPA 242
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      120  GVHYRYVADSELSSEE-----KKQLVYDIPTVVENDDTYILV----- 157
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

QY      243  NKSENLQSLKELYDPSAQRYSESGLVDFDPAKTIISRTPCNVAIPHGDHYHFTPYSKLS 302
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      158  -TKLSNQNLAEI---PNTGSKNEKQALVAG-----ASLAALGILI-----FAVSKKK 201
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

QY      303  ALEEKIARMPISGTCS--TVSTNAKPNEVSSGLSSNPSSLTTSKELSSADGYIFN 360
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      202  VKNKTVLHLVLVAGMGVGLVSVHLENELL-----LVNV-----TDVELSGEK--LPL 249
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

QY      361  PKDIVEETATATVLRGDDHFHYIPKSNQIGQPTPNNSLATPSPSLPNTGTSHEKHEED 420
      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,031A
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/861,804
FILING DATE: 21-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/SE91/00707
FILING DATE: 22-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9003374-7
FILING DATE: 22-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 012889-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1183 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-447-031A-2

Query Match 3.9%; Score 205.5; DB 2; Length 1183;
Best Local Similarity 19.7%; Pred. No. 3.6e-06;
Matches 214; Conservative 139; Mismatches 411; Indels 323; Gaps 50;

QY 16 NNRVYVDSGSSQSKENLT-PCQVSKQEGIAQIVIKITDQGYVTSNGHYHYNGK- 73
DB 195 NNRVYVDSGSSQSKENLT-PCQVSKQEGIAQIVIKITDQGYVTSNGHYHYNGK- 73
QY 74 --VPYDALFSEELMKDPNQLKADI-----VNEVKGYYLIKVDG-----KYVYVYLKRA 122
DB 239 AITDFEAFPGSKIVTONTNITDVTIPQGYSGYNSFSINIKYKTIENEOKEFVNNSQHW 298
QY 123 HADNVTKDEINROKQEHVDNEKVSNN-----AVARSQGRYTTN 163
DB 259 YQEH--GKEEVNGKSFNHTVHNANAGIEGTVKGLKVLKQDKTKAPIANVYKPLSKK 356
QY 164 DGYVF---NPADLIETGNAYIVPHGHHYIPKSDLSASLAAKAHLAGKNQPSOL 219
DB 357 DGVVKNQKQIEIITDANGI-----ANKALPSGDYILKEIEAPRYTFDRKE--- 406
QY 220 SYSSTADNNTOSVAKGSTSKPANKSENLOSLLKELYDSPSAQYSESGLVDFPAKILS 279
DB 407 -YPETMKDIDNQGVI--TTIENAKAIEKTDV-----SAQK-----VWEGTKVKV 448
QY 280 RTPNGVALPHGDHVFHPIYSK--LSALEEKIARMVPIGSGIVSTVSTNAKPNEVWSLSGL 337
DB 449 PTTFKLYKQDDNGNTTPVDAEIKKIED-----GTTKVTWSNLPEND----- 491
QY 338 SSNPSSLTTELKSSASDGYIFNPKDIVEATAYIVRHGDHFHPIKPS-NGIQOPTLP- 395
DB 492 -----EKVKT-----LDVTSETNMKY-----EFKDLPKYDEGKTEYTVT 514
QY 396 -----NNSLATPSPSLPI-NPGTSHEKHEEDGYCFDANRIIAEDSGFVMSHGDHNYTF 449
DB 515 GYTKENGLVVTNTEKTEITSISGEKVDKQDNQDKGR--PEKVSVNLLANG----- 565
QY 450 KDLTEEQIKAAQKHLEEVKTSNGHLSLSSHEQDYPGNKAKEMKOLDKKIE-EKTAGIMK 508
DB 566 -----EKVKT-----LDVTSETNMKY-----EFKDLPKYDEGKTEYTVT 600
QY 509 QYGVK-----RESIVYVK-----EKNALIPYHGDHHAADPIDEHKEVGVGISHSNYELFK 558
DB 601 EDHVKDYTTDINGTTINKTTPGTSATVTKNWDNNNO--DGKRTEI-----KVELY- 652

QY 559 PEGVA-----KREGNKV-YTGEELTNVYLLNLTSTFNQNF 595
DB 653 -QDKGATGKTALINESNNWTHWTGLDEKAGQOVYTVBELTKV-----KGYTHVDND 707
QY 596 LANGQKRVSPPELEKKLGINMLVKLITPDGKVLKYSKGVFEGE-----VGNIAN 648
DB 708 M--GNLIVTNKIPETTSISGEKVDKQDNQDKRREKVSVNLLADGKVKKLDVTSETN 765
QY 649 FELDQPYLP-----GQTFKYTIAS-----KDYPEVSYDGTFTVPTSLAYKMASQTFYFPHAG 701
DB 766 WKYEFKDLPKYDEGKTEYTVEDHVKDY--TTDINGT-----TITNKYTPG 810
QY 702 DTYLKV-----NPQFAVPKGTDALVRVDE-----FHGNAYLENNYKV 739
DB 811 ETSATVTKNWDNNQDKRPTKVELIYODGKATGKTALINESNNWTHWTGLDEKAG 870
QY 740 GEIKLPIPLNQ---GPTRTAGNKIPVTFMANAYLDNQSYIIVEFLEKENOTDK---- 792
DB 871 QQVXYTVBELTKVGYTHVDNNDMGNLIVTNKYTPETTSISGEKVDKQDNQDKRREK 930
QY 793 --PSILPQFKENKA-----OENSKLDE--KVEEPTSEKVEKELSETGNT 835
DB 931 VSVNLLANGKVKTLDTSETNWKYEFKDLPKYDEGKTEYTVEDHVKDYTTDINGTTI 990
QY 836 SNSTL--EEVPTVD-----PVOEKVAKFAESYGMKLENVLFNMDGTIELYLP 880
DB 991 TNKYTPGTSATVTKNWDNNQDKRPTKVELIYODGKATGKTALINESNNWTHWTG 1050
QY 881 SGEYIKKNMADFT--GEAPOGNGENKPSKSVSTGVNENQPTENKPADSL--PEAPNEK-- 936
DB 1051 LDEKAGQOVYTVDELTKVNGYTHVDNNDMGNLIVTNKYTPKPKRPIYKPEKPKDTP 1110
QY 937 FVKPENS 943
DB 1111 PTKPDHS 1117
RESULT 8
US-08-296-791-6
Sequence 6, Application US/08296791
Patent No. 524537
GENERAL INFORMATION:
APPLICANT: St. Gene III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
US-08-296-791-6

```

Query Match 3.8%; Score 202; DB 4; Length 1848;
Best Local Similarity 19.3%; Pred. NO. 1.3e-05;
Matches 233; Conservative 130; Mismatches 397; Indels 448; Gaps 55;

QY	111	DGKYVYVLDAAHADNVR--TKDEINRQKE---HYKDNEKVNNSVA-----VARSGOR	150
DB	132	ENRYITVEKNFTENVITSEKKEQQAQRREDDYVPRLDKFEVTEPATEASTANNKKE	191
QY	160	YTTNDGVVFNPA-----PAFVRLGSGTQFYKKGSRVQLILTEKQGNLLNRWDVGGDNLELVNA	180
DB	192	YNNSDKY---PAFVRLGSGTQFYKKGSRVQLILTEKQGNLLNRWDVGGDNLELVNA	248
QY	181	YIVPHGG-----HYHYIPKSDLSASELA-----DIIEDTGN	203
DB	249	YTYGIAGTPKYVHNENNGLIGFONSKREEHSDPKGLISQDPLTNAYVLGDSGFLEVDRE	308
QY	204	-----AAKAHLAGKNMOPSQ-----LSYSS--TASDNNNTQSV--AKGSTS	239
DB	309	KGKWLFLGSDFWAGYNKRSQWENIYKHEFAEKIYQOYSAGSLGTSNTOYTQWQATGSTS	368
QY	240	KPANKSENLOSILKELYDSPAORY-----SESQGLVDPDAKILISRTPN	283
DB	369	TIRGGGEPLSVDITDCKDPNHGKSTILKSGSCTLTLLNNHIDGAGLFEFGDYEVKGTSID	428
QY	284	-----GVAIPHGD-----HYHPIPYSKLS-----ALEE	306
DB	429	STTWKGAHSVADGKVTWKVHNPKYDLAKIGTKGLTVVGEKGKNEGLLKVGDTGTVILQ	488
QY	307	K-----IARWPISTGISTVSTNAK---PNEVVSLL--GSLSNPPSSL-----	344
DB	489	KADANNKQVAFSGVIGSVGRSTLVLDNDXQVDPNSITFYFGRGRDLGNSLTDPHIRNI	548
QY	345	-----TTSKLSASDCYIENPKDIVEEATATAYIVRHGHDFI-----	381
DB	549	DGGAHVYVHNWNTSNTITGESLTNP-----NITSYNIEAQODDPLRISRTPIYQOL	603
QY	382	YIPKSNQIOGTPPNNSLATPPSPSLPINTGTSHEKHEEDGYGFDA-----NRIIADES	435
DB	604	YFNQDNR--SYITLKKG--ASTRSELPQNSGESNENWLYMGRTSDAAKRVNHNHINERN	660
QY	436	GFVNSHGDHNNHYEFKKDLTEEOIKAAQKHLBEVKTSHNGLDLSLSHEODY--PCNAKEMK	493
DB	661	GF-----NGYF-----GEETKATQN--GKLNVTYFNG-----KSDQRFLTLTGSTNLNG	702
QY	494	DLDKKIEKTAGIMKOYGVKRSIVVYVNEKNAIITYPHGDHHHADPIDEHKPVGIG-----	548
DB	703	DL-----NVEKGTGLFSLGSRPTPHARDI-----AGISSTKKD	733
QY	549	-HSHSNVELPKPEGVAKK-----EGN-KVYTGEELTWVYNLLKNTSTNNQNFPLAN	598
DB	734	PHTEENNVVEDDNTNRFKATTMTVGTNAGLSIGR---NVANKITSITASN-----N	784
QY	599	GQKRVSFPPELEKKLGINMLVK-----LITPDCKVLEKVSXGVFEGV-GNI-----	646
DB	785	AQVHIGY-----KTGDTVCVRSDYTGVTCHNSNLSEKALNSFNFTLNRGNVULTEN	836
QY	647	ANFELQOYLPQGTQFYKTTTASDKDYFEVSDGTFTVPTSLAYKMASQITTFYPHPAGDIYLR	706
DB	837	ASSTLQKANLFG-----TTSIGTGSOVN-----LKSNHSHWLTGNSNVNOLNTNGHIH	885
QY	707	VNPQFAPVPGXTDALVRVDFEFGNA-----VLENNKYGEIKLPT-----	746
DB	886	LNAQDANKVTTYNTLTVNSLSGSGSFYVWDFNTNKNKVVVVKNSATGNTFTLOVADKGTG	945
QY	747	-PKLNGOT---TRTAGNKIPIYTFMANAYLDN-----QSTYIYVEVILEKENGT-	790
DB	946	EPNHNELTLFDASNATRNRLNLEVTLANGSVDRGAKYKILRVNNGYDIYLNPEVEKRNTO	1004

```

791  -----DKESILLPQFKRNKAQENSKLDEKVEEP----- 817
Db 1005 DTTNITTPNDIQADAPS-----AQSNNEIARVETPVPPAPATSAIASEQPETR 1055
QY 818 -----KTSKVEKEKILSETGNSTNSTLEEYPTVDPVOEKVAKFAESYGMKLE 865
Db 1056 PAETAOFAMETTNASTITAPKSDTATOTENPNSESVP-----ETTEKVAENPPQONE 1110
QY 866 NVLFNMDGTIELYPSGEVKKNMADF-----TGAPOCGNGENKPSNGSKYSTGVENQP 920
Db 1111 TVAKNECEATEPPONGEVAKEQDPVEANTQTNATSEKGTSTEQ-----TRETKESEP 1165
QY 921 T-----ENKPADSLPEAPNEKPV-----KPNSTNDNGMLNPEGVNGSDP-MLDPALAEA 968
Db 1166 TESVTSVNGPKVTSQSTEDKVVEKEEKAKVETEETQKAPQVTSKEPPKQAPAEV 1225
QY 969 PAVDPQOE 976
Db 1226 PTDNREE 1233

RESULT 9
PCT-US95-10661A-6
: Sequence 6, Application PC/TUS9510661A
: GENERAL INFORMATION:
: APPLICANT: Washington University, et al.
: TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/10661A
: FILING DATE: 16-AUG-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/296,791
: FILING DATE: 25-AUG-1994
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Treccartin, Richard F.
: REGISTRATION NUMBER: 31,801
: REFERENCE/DOCKET NUMBER: FP-59941/RFT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1848 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: PCT-US95-10661A-6

```

Query Match 3.8%; Score 202; DB 5; Length 1848;
Best Local Similarity 19.3%; Pred. No. 1.3e-05;
Matches 233; Conservative 130; Mismatches 397; Indels 448; Gaps 55;

```

QY 111 DGYYVYLKDAAHADNVR--TKDEINRQOE---HYKDRKYSNVA-----VARSOGR 159
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 132 ENRYIYVEKNFFETENVTSFKREEQDAQRREDDYMRPLDKFVTEVAPIEASTANNKGE 191
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 160 YTNDGCVYFNPA-----DIEDTGN 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 192 YNNSDKY---PAFVRLSGSGTQFIYKKGSRVQLILTEKDKQGNLLRNMDVGDGNLELVGNA 248
Qy 181 YIVPIGG-----HYHPIKSDLSASELA----- 203
Db 249 YTYIAGTPTKYVNHENGLIOFGNSKEHSDPKGILSDOPLTNVAVLGDSGSPLEFVVDRE 308
Qy 204 -----AAKHLAKNNQPSQ-----LSYSS-TASDNNQTSV--AKGSTS 239
Db 309 KGKWLFLSGYDFWAGYNNKSWQENNIYKHEFAEKIYQOYSAGSLIGSNQTYQWQATGSTS 368
Qy 240 KPAENSENLQSLKELYDPSAQR-----SESDGLVDFDAKILISPTN 283
Db 369 TITGGEPLSDVLDGDKPNHGKSIITUGSGTLTLNHNIDOGAGGLFFEGDYEVKGTSD 428
Qy 284 -----GVAIPHGD-----HYHPIYSKLS-----ALEE 306
Db 429 STTWKAGVSVADGKTVTKVHNPKYDRLAKIGKTLVVEGKNGEGLKVGDDGTIVLKQ 488
Qy 307 K-----IARVPISGTGVSTNAK---PNEVWSSL--GSLSSNPSSL----- 344
Db 489 KADANKVQAFSQGVIVSGRSTVLNDDKQVDPNSIYFCFRGRLDLNGLSTFDHIANI 548
Qy 345 -----TTSKELSSADGYIFNPKDIVEETATAYIVRHGDHFF----- 381
Db 549 DGARVNHNTNITNITGESLTNP-----NTITSYNIEAQDDHPLRISPIYQOL 603
Qy 382 YLPKSNQIGQPTLPNNSLATPSPINPGTSHEKHEEDGYGFDA-----NRIIAEDS 435
Db 604 YFNODNR-SYYTLKKG--ASTRSELSPONGSESNWLYMGRTSDAAKRNVMNHINERNM 660
Qy 436 GPMVSHGDHNYFFKKDLTEEDIKAAQKHLEVKTSNGLDLSLSHEQDY--PGNAKEMK 493
Db 661 GF-----NGYF-----GEEETKATON--GKLVNFTFG-----KSDQRFLLTGCTNLNG 702
Qy 494 DLDDKIEEKIAGIMKQYGVKRESIVVNNREKNAIYIPHGDDHHADPIDHKKPVGIG----- 548
Db 703 DL-----NVEKGTFLSLGRPTPHARDI-----AGISSTKDD 733
Qy 549 -HSHSNYELFKPEEGVAKK-----TGN-KVYTGEBELTNVNNLLKNSFTNNQNTFLAN 598
Db 734 PIHTENNEVVVDDINRNFKATTNVTGNASLYSGR--NVANITSNITASN-----N 784
Qy 599 GOKRVSFSPPELEKKLGINMLVK-----LITPDGKVLKVGSKVFGGV--GNI----- 646
Db 785 AQVHIGY-----KTGDTVCVRSYDTGYTCHNSLSEKALNSFNPTNLGRNVNLJEN 836
Qy 647 ANFELDQPLPGOTKYTTASKDYDEVSVDGTFVTPTSLAYKMASQTFYFPFHAGDTYLR 706
Db 837 ASFTLGAHLFG-----TIOSGTGSOVN-----LKNSHMLTGNNSNVNOLNLTNGHIH 885
Qy 707 VNPQFAPVPGTDLAVRVDFEFGNA-----YLENNYKVGELKLPIT----- 746
Db 886 LNAQDANKVITYNTLTVNSLSNGSGFYVWDFTNKSNKVVVNSATGNTFLQVADXTG 945
Qy 747 -PKLNGOT---TRTAGNKIPVTFFMANAYLDN-----OSTYIYEVPILEKENOT- 790
Db 946 EPNHNELTLFDASNATRNRLVET--LANGSVDRGAWKYKLRNVNGRYDLYNPEVEKRNQTV 1004
Qy 791 -----DKPSILPQFKRKAQNSKLDKVEEP----- 817
Db 1005 DITNITTPNDIQADAPS-----AQSNNEEIARVETPVPPPPAPATSAIASQEPETR 1055
Qy 818 -----KTSEKVEKELSETGNSSTLSLEVTVDVDPQVKVAKFAESYGMKLE 865
Db 1056 PAETAQAMEETNTANSTAPKSDTATQTPNPSVESVPS-----ETTEKVAENPPQENE 1110
Qy 866 NVLFNMDGTIELYLPSEGVYKKNWADF-----TGEAPQNGENKSPENKYSTGTGVENOP 920
Db 1111 TVAKNEQATEPTPQNGEYAKEDQPTVEANTQTNATSEGTETEQT-----TATKSEP 1165
Qy 921 T-----ENKPADSLPEAPNERPV-----KPENSTDNGMLNPEGVNGSDP--MLDPALEEA 968
Db 1166 TESVTSYSENOPEKTVSQSTEDKVVVEKEKAKVETEETQKAPQVTSKEPPKQAPAEPEV 1225

Qy 969 PAVDPVQGE 976
Db 1226 PTDTNAEE 1233

RESULT 10
US-08-961-083-4
; Sequence 4, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 571 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-4

Query Match 3.8%; Score 200; DB 4; Length 571;

Best Local Similarity 22.6%; Pred. No. 3e-06;
Matches 166; Conservative 87; Mismatches 222; Indels 258; Gaps 42;

Qy 361 PKDIVEETATAYIVRHGDHFFHYIPKSNQIGQPTLPNNSLATPSPSLPINPCTSHKHEED 420
Db 17 PRELSGTYTYGYI-KEGTTTSEVSNO-----KSSVATPT-----KQKV 56
Qy 421 GYGFDANRIIAEDSGFVMSHGDHNYFFKKDLTEEDIGKAAQKHLEVKTSNGLDLSLS 480
Db 57 DYNVTPNFV---DHPSTVQAIQEQTPVSSTKP-TEVQVWEKPFSTELINPRKEEKSSDS 112
Qy 481 HEQDYPGNAKEMKDLKKIEKTAGIMKQYGVKRESIVVNNKNAIYIPHGDDHHADPID 540
Db 113 QEQ-----LAHRNLETKKEEKIS-----PREKTGV-----NTLNPD 145
Qy 541 E-----HKPVGIGHSHSNVELFKPEEGVAKKCKNKKVYTGEBELTNVNNLLKNSFTNNQNF 594
Db 146 EVLSGLNKP-----ELLYREETME---TKIDFOEEL-----QENP 178
Qy 595 TLANGQRKRVSEFPPELEKKLGINM-LVKLIT-----PDGKVLKVS 636
Db 179 DLAEGTVRV-----KQEGKLGKKEIVRIFSVNKEEVSREIVSTSTAPSPRIVEKGT 232

```
QY 637 K--VFGE-----GV-----GNANFELDQYLPQGTFKYTIASKDYPEVSYDGTFTVPT 693
Db 233 KTVIKQPTGVHEKDVQGAIVEPAI-QPELP-----EAVSDKGEPEVQP-----TLPE 283
QY 684 SLAYKMASQTIFFPHFAGDTYLRVNPQ-----FAVPKGTDALVRVDFEFHGNAYLENNYKV 739
Db 284 AVV-----TDKGET--EVQPESDPTVVSVDKGEPEQVAPLPEYKGN-----I 322
QY 740 GEIK--LPIPKL--NOGTRTAGNKKIPVTFMANAYLONQSTIYVEVILKSNQTDKPSIL 796
Db 323 EQVRPEFVEKTEQGEPEKT--EEVPV-----KPTETPVNNEGTEGTSI- 367
QY 797 PQFERN---KAQNSKLDEKVEEPKISEK-----VEKEKLSETGNST 835
Db 368 -QEAENPVQPAESTINSEKV-SPDTSSKNTGEVSNPNPDSSTISVCSNKPENHNSKNEN 425
QY 836 SNTLEVPVTPVQVKVAKFAESYGMKLENVLFNMDGTIELYLPGEVIKKNMADFTGE 895
Db 426 SEKTVEVP--VNP-----NEGTFE-----GTSNQETE 451
QY 896 APOGGENKPSNGKV---STGTVENOPTENKP--ADSLPEAPNEKPVKPKENS-----TDN 946
Db 452 KPQPAEETQTSKGTANENTGVSKNPKSDSKPPVESNOPEKNGTATPENSNGTTSN 511
QY 947 GMLNPE---GNVGSDDPMLDPALEAPVDPVQEKLEKFTASYGLGLDSVIFNMDGTIELR 1003
Db 512 GQTEPEPSNGNSTEDVSTESNTSNGNEIKOENE-----LDPDKKVEPEKTELELR 564
QY 1004 LPSSEVIKKNLSD 1016
Db 565 -----NVSD 568
```

RESULT 11

```
US-09-071-035-326
: Sequence 326, Application US/09071035
: Patent No 6448043
: GENERAL INFORMATION:
: APPLICANT: Gil H. Choi
: TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
: NUMBER OF SEQUENCES: 496
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/071,035
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: A. Anders Brookes
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB369P2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 326:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 851 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
```

```
: MOLECULE TYPE: protein
US-09-071-035-326
```

```
Query Match 3.7%; Score 198; DB 4; Length 851;
Best Local Similarity 19.2%; Pred. No. 7.6e-06;
Matches 190; Conservative 114; Mismatches 384; Indels 302; Gaps 40;

QY 14 KONNRVSYVDGS-----OSSQKSENLTDPQVSKQEGIOAEQIVIKITDQGYVISHGDHYH 68
Db 74 EDNQLRVQOEAVPSVLKSVSSGDLYPD-----GSLIOHG----- 109
QY 69 YNKGVPYDALFSEELL-----MKDPNYQLKDADIVNEVGGYI--IKVD 111
Db 110 YF-----PINGVGNELLKGFGRIOITLOGSDWEMDPNI-----SNLFNVVDKGYLDLMVN 161
QY 112 GRYVYVLKDAAHADAVRTKDEINROKQEHVKDNEKVNNSVAVARSGQRYTTNDGVVENPA 171
Db 162 GK---MPSMVSGRSISRAPETNPFTTEFESGKETIANLTLIAKFAPELNLDIYT---- 213
QY 172 DLIETGNAYIVPHGGHYHYIPKS-----DLSASELAARAKAHLAKGNMOPSQLSYST 224
Db 214 -----SIOTWLOQSGSYHFFKXKPRDFEALIDLKXVNSASPAQATPMOSLNIVYGSMDRV 268
QY 225 ASDNNTQSAKGTSTKPA-----NKSENLSL-----LKELYDSPAQRYSESGLVDFDP 274
Db 269 LOKNNEYAVGISMSYRVGNVYEFNGTENKKGWHTADGMLYLYNQDFAQ--FDEGYWATIDP 327
QY 275 AKLIISRTNGVAIPHGDHYHFTPYSKLSALEEKIARMVPISTGTSTVSTNAKPNEVWSSL 334
Db 328 YRLPGTTVDTRLANGAY-----TGKRSPQSVWVGS-----NNGQVASI 366
QY 335 QSL--SSNPSSLTTSKELSSASDGVIPNPKDIVEETATAYIVRHGDHPHYIPKSNQIQOP 392
Db 367 GMFLDKSNEGMLVAKKSWFLDGOIINLGSITTTDASTIETILDNRMIHPQEVKLMQ 426
QY 393 TLPNNSLATPSPSLPINPGTSHEKHEEDGYGFDANR-----IIAEDSGFVMSHGDHNYF 448
Db 427 SDKDNSWISLSAAXPLN-----NTGYVFPNSMNTLDVQIERSG---RYGDINEYF 474
QY 449 FKDDITEQIRAAQKHLBEVKTSHNGLSLSHEDGYPGNAKEMKDLKKEEAGTMR 508
Db 475 VN-----DKTYTNTFAKISK 489
QY 509 QYGVKRESIVVNKEXNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELF-----KPEEVA 564
Db 490 NYGKTVE-----NGTYEYLTVVGXTNEDIA 514
QY 565 KKEGNKYVTGCELINVVNLLKSNFNNOFTLANGCKKRVSFSPPELEKKGICINMLVKLI 624
Db 515 ALSKNKGYT--VLENTANLQAEAGNYMMNTWNNNDQEIAGLAYD----- 558
QY 625 TPDGKVLKSVSKVGEVGVGN-----IANFELDQYLPQGTFKYTIASKDYPEVSYDGT 678
Db 559 -PMSVISEKIDNGVYRLTANPLQNNASVSTIEFDKCIL-----EVVAAD--PEISVDQN 609
QY 679 FTVPTSLAYKMASQTIFFPHFAGDTYLRVNPQFAVPKGTDALVRVDFEFHGNAYLENNYK 738
Db 610 IITLNSAGLNGSSRSII-----VKTTPEVT-----KEALEKLIQF--QKEHQEKDYT 654
QY 739 VGEIKLPIPKLNOGTT---RTAGNKIPVTFMANAYLONQSTIYVEVILKSNQTDKPSI 795
Db 655 ASSWKVYSEALKQAQTVADQITATQAEVD--QAETELNAVQLVKVPYTK-----VDKTNL 709
QY 796 LPQKRNKAQNSKLDEKVEEPKTSSEKVEKEKLSSETGNSTSELEEVPTVDPVQEVAK 855
Db 710 LKIITK-----ENKHEQEK--DYTASSWKVYSEALKQ--ACTVADQITATQAEVDQAEAKLRS 762
QY 856 FAEYGMKLENVLFNMDGTIELYLPSEGEVIKKNMADFTGEAPQNGENKPKSEN---GKVS 912
Db 763 AVKRLTLK-----NSGENKKE-----QKNGNGNHLNSTGTVDQ 796
QY 913 TGTVENQPT-----NKPADSLPEAPNEKPV 938
```

Db 797 TGTQVQKPSOGGFRKASQFLPSTGEKXSI 826

RESULT 12

US-09-071-035-330

; Sequence 330, Application US/09071035

; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/071.035

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

Query Match 3.7%, Score 198, DB 4, Length 851:

Best Local Similarity 19.2%, Pred. No. 7.6e-06;

Matches 190; Conservative 114; Mismatches 384; Indels 302; Gaps 40;

QY 14 KDNRRVSVYDGS-----QSSOKSENLPDQVQSKGEGIAQIVIKITDQGYVTSBGDHYH 68

Db 74 EDNQLVQVQEAQSVLKSQSGGLFPD-----GSLIQHG----- 109

QY 69 YNKGVPYDALFSEBELL-----MKDPYQKADIVNEVKGYYI-IKVD 111

Db 110 YP-----PYNGSYGNELKGFGRIGTILQSGDWEMNDPNI-----SNLFNVYDKGYLQLMVN 161

QY 112 GKYYVYLKAAHADNVRTKDEINQKQEHVKDNEKVNVAQVPSQGRYTTNDGYVFNPA 171

Db 162 GK-----MFSMVSGRSISAPETNFTTEFESGKETIANLTIKAFAPENLNDIYT----- 213

QY 172 DIIETGNAYIVPGHHYHYTPKS-----DLSASELAHAHLAGKNQPSQSYSS 224

Db 214 -----SIQTLWQQSGSYHYHFKPKRDFEALIDKNVYNASPAQATPMQSLNLYGSMDRV 268

QY 225 ASDNNTQSVAKGTSKPA-----NKSENLOS-----LKELYDSPAQRYSQSDGLVEDP 274

Db 269 LQKNNEYAVGISMYSORVNGYFNGTENKKGWHTADGMLYLNQDFAQ-PDEGYWATIDP 327

QY 275 AKIISRTNGVAIPGHDHYHPIPYSKLSALEEKTARKVPISTGTGTVSNVNAKNEVVSSL 334

Db 328 YRLPGTVDYTRLANGAY-----TGKRSQSWYGG-----NNGQVASI 366

QY 335 GSL--SSNPSSLTTSKELSSASDGYIENPKDIVETATAYIVRHGDHFHYIPKSNQIGQP 392

Db 367 GMFLDKSNEGNNLVAKKSWFLDQIILNLSGITGTITDASIEITLDNRMTHPQEVKLNQ 426

QY 393 TLPNNSLATPSPSLPIPGTSGHEKHEEDGFGFANR-----IIAEDSGGFVMSHGDNHIF 448

Db 427 SKDNKSWISLSAAAXPLN-----NIGYVFPNSMWTLDVQIEERSG-----RYGDLNIEYF 474

QY 449 PKDLTEBQIKAAQKHLEEVKTSNGLDLSLSSHDQYPGNKAKEMKDKKIEBKIAKIMK 508

Db 475 VN-----DKTYTTFAKISK 489

QY 509 QYGVKRESIVVKNKNAIIPHGDDHHPIDEBKPVGIGHSHSNYELF-----KPEGYA 564

Db 490 NYGRTVE-----NGTYEYITVVGKTNEEIA 514

QY 565 KKEGNKVTGEBELTNVNNLLKNSTFNQNTFLANGQKRVSFSPPELEKKGINMLVKLI 624

Db 515 ALSKNKGYT--VLENTANLOAIEAGNVVMKNTWNNDOEIAGLVAYD----- 558

QY 625 TPDGKVLKSVKVGEGVGN-----IANFELDQVLPQGTQYTIASKDYDEVSYDGT 578

Db 559 -PMSVISEKIDNGYRILTLANPLONNASVSEIEFDKGL-----EWAAD-PEISVQON 609

QY 679 FTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAVPRKGTDALYRVFDEFHGNAYLENNYK 738

Db 610 IITLNSAGLNGSSRSII-----VKTTPVET---KEALEKIQE--QKEHQEKDYT 654

QY 739 VGEIKLPIPKLNQGT---RTAGNKIPVTMANAYLDNQSTYIVVPILEKENCTDKPSI 795

Db 655 ASSWKVYSEALKQAQTVADQTTATQAEVD-QAETELSAVKQLVKVPTKE---VDKTL 709

QY 796 LPQKRKAQENSKLDKVEPKTSEKVEKEKISETNSNSTLEEVTVDPVQEKVAK 855

Db 710 LKLIK-----ENEHQEK-DYTASSWKVYSEALKQ-AQTVADQTTATQAEVDQAEAKLS 762

QY 856 FAESYGMKLENVLFNMDGTTIELYLPSCGEVYKKNMADTGEAPOGNGENKPSN---GKVS 912

Db 763 AVKRLTLK-----NSGENKKE-----QKNGNGHNLNTSTGVQDQ 796

QY 913 TGTVENQPTQ-----NKPADSLPEAPNEKPV 938

Db 797 TGTQVQKPSOGGFRKASQFLPSTGEKXSI 826

RESULT 13

US-09-071-035-334

; Sequence 334, Application US/09071035,

; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/071.035

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 334:
SEQUENCE CHARACTERISTICS:
LENGTH: 851 amino acids
TYPE: amino acids
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-334

Query Match 3.7%; Score 198; DB 4; Length 851;

Best Local Similarity 19.2%; Pred. No. 7.6e-06;

Matches 190; Conservative 114; Mismatches 384; Indels 302; Gaps 40;

QY 14 KDNRRVSYVDGS-----QSSQKSENLTDPQVSQKEGIAEQIVIKITDQGYVTSHGDIYH 68
Db 74 EDNQLVQVEAVPSVLKSVSGDGLYD-----GSLIQHG----- 109
QY 69 YXNGKVPYDALSEELL-----MKDPYQLKADADIVNEVKGYYI-IKVD 111
Db 110 YF-----PYNGSYGNELKLGFGRIQITLOGSDWEMNDPNI-----SNLFNVYDKGYLQLMVN 161
QY 112 GXYVYLKDAAHADNVRTKDEINROKQEHVKDNEKVNNAVARSQGRYTTNDGVVFNP 171
Db 162 GK-----MPSWVSGRSIRAPENPTTFESGKETIANLTLIAKAPENLRNDIYT----- 213
QY 172 DIETGTNAYIVPHGHHYIYKPS-----DLSASELAACHAGLAKNMOPQSLSYST 224
Db 214 -----SIQTLOQSSYVHFKKPRDFEALIDLKNVNASAPAAQATPMQSLNVYGMDRV 268
QY 225 ASDNNTQSVAKGSTSKPA-----NKSLENQSL-----LKEYDPSAORYSESGLVDP 274
Db 269 LOKNEYAVGISMYQORVNGVEFGNTENKKGWHTADGMLYLNQDFPAO-FDEGVWATIDP 327
QY 275 AKIISRTNGVAIPGDHYYHFTIPYSKLSALEBKIAKMYVPISGTGSTVSTNAKPNEVSSL 334
Db 328 YRLPGITVDITRELANGY-----TGKRSPOSQSVWGS-----NNGQVASI 366
QY 335 GSL--SSPSSLTTSKELSSASDGYIENPKDITVEETAYIVRHGCHDHPHYIPKSNQIQP 392
Db 367 GMFLDKSNEGMNLVAKKSWFLDGOIIMLGSGITGTTDASITILDNMIHPQEVKLNQ 426
QY 393 TLPNNSLATPSPSLPINPGTSHEKHEEDGYGFANDR-----IIAEDSGFVMSHGDHNYF 448
Db 427 SOKNSWISLSAAXPLN-----NIGYVFPNSMNTLDVQIERSG---RYGDINEYF 474
QY 449 FKKDLTEQIRAAOKHLEEVKTSNGLDLSLSHEDQYPCNAKEMKDKKTEEKIAGTMK 508
Db 475 VN-----DKTYTNTFTAKISK 489
QY 509 QYGVKRESIVVNEKKNATIIYPHGDHHDADIDEHKPVGIGHSHSNYELF-----KPEQVA 564
Db 490 NYGKTVE-----NGTYEYLVTVGKTNEEIA 514
QY 565 KKEGKVVITGELTNVWLLKNSFTNNQNTFLANGQKRVSPFSPEPEKKGILNMLVKLI 624
Db 515 ALSKNKGT--VLENTANLQAEAGNYVMNTWNNDQEIAGLAYD----- 558
QY 625 TPDGKVLKSVKVFGEVGN-----IANFELDQYLPQGTETTYTASKDYPEVSYDGT 678
Db 559 -PMSVISEKIDNGVYRLTLANPLQNNASVSEFPKGI-----EVVAAD-PEISVDQN 609
QY 679 FTVPISLAYKMASQITVYFFHAGDTYLRVNFQFVAPVKGTDALVRVDFDGHGNAYLENNYK 738
Db 610 IITLNASAGINGSRSII-----VKTTPVET-----KEALEKLQIE--QKEHQENDYT 654
QY 739 VGEIKLPKLNQGT---RTAGNKIPYTFMANAYLDNQSYTYVEVPILEKENGTDRPSI 795

Db 655 ASSWKVYSEALKQAQTADQTATQAQVVD-QAETELRSVAVQLYKVPKTE-----VDKTNL 709
QY 796 LPQFKRNKAQENSKLDEKVEEPKTSKVEKEKLSGTGNTSTNSLLEEVPTVDPVQEKVAK 855
Db 710 LKIIK-----ENERHOEK-DVTASSWKVYSEALKQ-AQTADQTATQAQVVDQAQAKLRS 762
QY 856 FAESYGMKLENVLFNMDGTIELYLPSEGEVIRKKNMADFTGEAPQNGENKPKSEN---GKVS 912
Db 763 AVKRLTTLK-----NSGENKKE-----QKNGGNHGLNTSTGVQDQ 796
QY 913 TQTVENQPT-----NKPADSLPEAPNEKPV 938
Db 797 TGTQVKPSSOGGFRKASQFLPSTGEKKSII 826
RESULT 14
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: LYON Douchette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080
Query Match 3.5%; Score 188.5; DB 4; Length 3696;
Best Local Similarity 18.8%; Pred. No. 0.00033;
Matches 204; Conservative 163; Mismatches 407; Indels 309; Gaps 49;
QY 13 KDNRRVSYVDGSQSS--QKSENLTDP-----QVSQKEGIAEQIVIKITDQGYVTSHG 65
Db 1911 NSDTER--EVNGAKTNGLEKINNIOPTQTNTAKQKINDKAQEQIQTINPTDPT---- 1964
QY 66 HYHYNGKVPYDALFSBELLMKDPYQLKADIVNEVKGYYIIVKDGKYYVYLKDAHAD 125
Db 1965 -----EE-----EKQEATNRYNAGLAQAIQ-----NIN 1987
QY 126 NVRTKDEINROKQEHVKDNEKVNNAVARSQGRYTTNDGVYFNPADIIETGNAYIVPH 185
Db 1988 NAHSTQEVNESKTSNIATIKRSVQPNVIKKPTAINSLTQEAN--NOKTLIGNDGA----- 2040
QY 186 GGHVHYIPKSDLSASELAACHAGLAKNMOPQSLSYSTASDNTQSV-AGST-----S 239
Db 2041 -----TDDEKAAKQLVTKLNEQIQIKIHST-QDNQVDNVKAAITA-KLIN 2087
QY 240 KPANKSENLSLLKELYDPSAORYSESGLVFDPAKIIISRTPNGVAIP-----HGDIHYF 295
Db 2088 ANAHKRODAINILTNLAESKKSDIRANQDATTETKNTAICSIDDTLAQARNNINGANTNA 2147
QY 296 IPYSKLSALEEKIARMYPISGTGSTVSTNAKPNEVSSLG-----SLSSNPSSLTTSKE-- 349
Db 2148 LVDLENEDGGKQKQIRIV-----LSTQTKIOAKA-DIAAIGQQQSTIDONQATTEKQEA 2202
QY 350 ---LSSASDGY-----IFNPKDIVEETATAYIVRHGDHFIYI-PKSNQI----- 389
Db 2203 LERLUNQETNGVNDRIQAAALANQNVTDKNNILET-----IRNVEPIVIVKPKANEIRKK 2257
QY 390 --GOPTLPNNSLATPSPSLPINPGTSHE-KHE---EDGYGFDANRIIAEDSGFVMSHG 443
Db 2258 AAEOTTLINQDQATLEEKQATGALKEEVKNEALNQVSAHSNNDVKIAENNGIAKISEV 2317

444 HHNYFFKDLTEQIKAAQKHLEEVKTSNGLSDLSLSSHEQDYPGNK-----490
Db 2318 HPETIIKRNKAEIQQDQASQIDTNNANKSTNEEKSAAIDRVNVAKIDAINNITNAITTT 2377
QY 491 -ENKDLKKIEKIKAGIMQYGVKRESIVV---NKEKNATIIYPGDHHHADPIDEHKPV 545
Db 2378 QLVNDKNSGNTSISQILPSTAVKTNALAAASEAKNNAIL-----DQTPNATAEEK--2430
QY 546 GIGHSHSNVELKPEGVAKKGN--KYVTGEELNV-----VNLKKNSTENN 591
Db 2431 ----EENKRVDELQ--EADANILKAHTTDEVANNIKNAQVONNAVQVEVIKKONVKN 2483
QY 592 ONFTLANGOKRVSFSP-PELEKKLGINLVKLITPDGKVLKSVKGVFEGGVNI-ANF 649
Db 2484 QLNQFDNOKKIILNTPDALLEKAEANRLQ-----NVLSTS-----DEIANVDHNN 2532
QY 650 ELDQYLPQGTFRKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRNP 709
Db 2533 EVDQ-----ALDRAPKIE-----EIVP 2550
QY 710 QFAYPKGTALYRVFDEFHGNAYLENNYKVGELKLPKLNQGTTRTAGNKIPVTFMANA 769
Db 2551 Q--VSKRDVNLAIQEA-----NSOTQELQ-----2574
QY 770 YLDNQ-STYIVEPILEKENQTDKPSILPQFRN--KAQENSKLD-----EKVEE 816
Db 2575 --ENQATNEEKTEALNKINQ-----LLNAQVNIQDQSNKVDSDAKTRSIODIEQIQP 2627
QY 817 PKTSEKVEKELSETGNSNSTLEVP--TVDPVQEKVAKFAE-----SYGM 862
Db 2628 HPQTKATGRHLRLEKANO-QOSTIATHPNSTIEEROEASAKLQELVKKAIAKIDKGQND 2686
QY 863 KLENVLFNMDGTTELPLSGEVIKKNAQFTGEAPOCNGENKPSNGKYSTGTVENQPT 922
Db 2687 DVKTVVNGTAEIENLIPATVYKAKADVNAEKEQN-----LQINSDEATTE 2736
QY 923 NK--PADSLPEAPNEKVPKPNSTNGMNLNPEGNGSDPMLD--PALEAP-AYDPVQEK 977
Db 2737 EKLVASDNLNHHVETTNQAEADPTNQVNEKNKGIGTIRDIQPLVVKKPTAKSKIESA 2796
QY 978 LEK 980
Db 2797 VEK 2799
RESULT 15
US-08-904-263A-4
Sequence 4, Application US/08904263A
Patent No. 6015889
GENERAL INFORMATION:
APPLICANT: LINDAHL, GUNNAR
APPLICANT: STALHAMMAR-CARLEWALM,
APPLICANT: MARGARETHA
APPLICANT: STENBERG, LARS
TITLE OF INVENTION: PROTEIN RIB, A CELL SURFACE PROTEIN THAT
CONFERS IMMUNITY TO MANY STRAINS OF THE GROUP B
TITLE OF INVENTION: STREPTOCOCCUS: PROCESS FOR PURIFICATION OF THE PROTEIN,
TITLE OF INVENTION: REAGENT KIT AND PHARMACEUTICAL COMPOSITION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

Query Match 3.5%; Score 186; DB 3; Length 1231;
Best Local Similarity 20.3%; Pred. No. 9.7e-05;
Matches 240; Conservative 145; Mismatches 416; Indels 384; Gaps 66;
QY 19 VSYVDGSSOSKSENLPDQVSQKEGIAEQIVIKITDQGVVTSHG---DHVHYNGKV 74
Db 35 ISFLGG--FTGQGFNISDVFRAEIVSGSAVTLN-TNMTKNVQNGRAYIDLVDKNGKI 91
QY 75 -PYDA-----LFSSELLMKDPNY--OLKDADIVNEVGKY-IKVDGKYY-----115
Db 92 DPLQILTLNSDPLKAQYVIRQGGNYFTQPSSELTIVGAASINVTVLKTDGSPHTKPDQQVD 151
QY 116 -----VYLKDAAH--ADNVRTK-----DEINRQK-----OEHVKNKENVNVA 152
Db 152 IINSLTIYNSAALRDKIDEVKKAEDPKWDEGSRDVL-SLDDIKTDIDNNPKTQSDIA 211
QY 153 VARSOGRTYNDGVFNPAIIEDTG-NAYIVPHGHHYIIPKSDLSASELAAKAHLAG 211
Db 212 -----NKITEVTNLEKILVPR-----IPDAD-----KNDPAG 238
QY 212 KMQPSQLSYSTASDNNQTSVAKGSTSKPANKSENLOSLLK---ELYDPSAQRYS 268
Db 239 KD-----QQVNVGETPKAEDSIGNLPLDKGTTVAFTP-VDTATPGD 280
QY 269 GLVFPDKAIIISRTPNG-----VAIPHGDHYHFIYPSKLSALEEKIARMVPISGTG 318
Db 281 ---KPAKVVVTPDGSKDTVDVIVKVVDPRTDADKNDP-----AGKD 319
QY 319 STVSTNAKPNNEVSSLSGSSNPSLTSKE--LSSASDG-----YTFNPKDIVEE 367
Db 320 QOVNVGETP-KAEDSIGNLPLDKGTTVAFTPVDATPGDKPAKVVVTPDGSKDIDV 378
QY 368 TATAYIVR-HGDHFIYIPKSNQI-----GOPTLPNNSLATPSPSLPINFETS 413
Db 379 TVKVVDPRTDADKNDPAGKQOVNVGETPKAEDSIGNLPLDKGTTVAFTPVDATP 436
QY 414 HKHEEDGYGFDANRIIADESGFVMSHGHNHFFKDLTEQIKAAQKHLEEVKTSN 473
Db 437 GDK-----PAKVVVTPDGS-----KDTVDVTVKVVDPRTDADKNDPA 474
QY 474 GLDSL-----SSHEODYPCNAKEMKDLDKKIE-----EKIAGIMKQY--GVK 513
Db 475 GKQOVNVGETPKAEDSIGN---LPDLKGTIVAFETPVDATPGDKPAKVVVTPDGS 531
QY 514 RESIVVNKEKNATIIYPHGDHHHADPIDSEKPVGIGHSHSNYELEKPEGVAKKGNK 573
Db 532 DTVDVTVK---VVDPTDADKNDPAGKQOVNVG-----ETPKAEDSIG-----572
QY 574 GELTNVNVNLLKNST--FNNQNFITLANGK--RVSEFSPPELEKKLGINMLVKL 624
Db 573 -----NLPLDKGTTVAFTPVDATPGDKPAKVVVTPD-----GSKDIDVTVKVVDPRTD 625
QY 625 ----TPDGKVLKSVSKV--FOEGVGNIANFELDQYLPQGTFRKYTIASKDYPEVSYDGT 678
Db 626 ADKNDPAGKQOVNVGETPKAEDSIGNL-----PDLPKGT---TVA-----663

APPLICATION NUMBER: US/08/904,263A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 552-119P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1231 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-904-263A-4

Search completed: May 13, 2003, 13:58:31
Job time : 36.7951 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 13:53:57 ; Search time 25.1868 seconds
(without alignments)
3723.142 Million cell updates/sec

Title: US-09-471-255-55
Perfect score: 5315
Sequence: 1 CAYALNHRQENKDNRRVS.....IELRLPGEVTKNLSDFIA 1019

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
1	5315	100.0	1039	US-09-884-465A-6
2	4322	81.3	1238	Sequence 6, Appl1
3	4276.5	80.5	840	Sequence 381, App
4	4261.5	80.2	1365	Sequence 10, Appl
5	4258.5	80.1	1378	Sequence 382, App
6	4237	79.7	1152	Sequence 378, App
7	3384	63.7	906	Sequence 379, App
8	3375	63.5	906	Sequence 370, App
9	3365	63.3	906	Sequence 332, App
10	3362	63.3	900	Sequence 372, App
11	3345.5	62.9	901	Sequence 371, App
12	3338	62.8	900	Sequence 334, App
13	3329	62.6	900	Sequence 343, App
14	3329	62.6	900	Sequence 335, App
15	3327.5	62.6	901	Sequence 339, App
16	3325	62.6	894	Sequence 341, App
17	3323.5	62.5	895	Sequence 345, App
18	3297	62.0	633	Sequence 337, App
19	3297	62.0	633	Sequence 347, App
				Sequence 350, App
				Sequence 354, App

ALIGNMENTS

RESULT 1

US-09-884-465A-6
Sequence 6, Application US/09884465A
Publication NO. US2003007293A1

GENERAL INFORMATION:

APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 1039
TYPE: PRT
ORGANISM: Streptococcus pneumoniae

Query Match 100.0%; Score 5315; DB 9; Length 1039;
Best Local Similarity 100.0%; Pred. No. 6e-302;
Matches 1019; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CAYALNHRQENKDNRRVSVDGSSQSKSENITPDQVSKEGIQAEQIVIKITDGYV	60
Db	21	CAYALNHRQENKDNRRVSVDGSSQSKSENITPDQVSKEGIQAEQIVIKITDGYV	80
QY	61	TSHGSHYHYNGKVPYDALFSEELMKDPNYQLKDADIVNEVKGYYIKVDGKYYVLKD	120
Db	81	TSHGSHYHYNGKVPYDALFSEELMKDPNYQLKDADIVNEVKGYYIKVDGKYYVLKD	140
QY	121	AAHADNVRTKDEINRQKQEHVKONEKNSNVAVARSGCRITTDNDGVFNPADIETGNA	180
Db	141	AAHADNVRTKDEINRQKQEHVKONEKNSNVAVARSGCRITTDNDGVFNPADIETGNA	200
QY	181	YIVPHGHHYHIFPKSDLSASELAALAKAGKNKQPSQLSYSTASDNNNTQSVAKGSTK	240

20	3297	62.0	633	9	US-09-884-465A-361	Sequence 361, App
21	3297	62.0	633	9	US-09-884-465A-362	Sequence 362, App
22	3288.5	61.9	895	9	US-09-884-465A-344	Sequence 344, App
23	3288	61.9	633	9	US-09-884-465A-351	Sequence 351, App
24	3288	61.9	633	9	US-09-884-465A-352	Sequence 352, App
25	3288	61.9	633	9	US-09-884-465A-357	Sequence 357, App
26	3288	61.9	633	9	US-09-884-465A-358	Sequence 358, App
27	3288	61.9	633	9	US-09-884-465A-359	Sequence 359, App
28	3288	61.9	633	9	US-09-884-465A-333	Sequence 333, App
29	3288	61.9	900	9	US-09-884-465A-360	Sequence 360, App
30	3288	61.9	906	9	US-09-884-465A-369	Sequence 369, App
31	3288	61.9	906	9	US-09-884-465A-373	Sequence 373, App
32	3287	61.8	633	9	US-09-884-465A-355	Sequence 355, App
33	3287	61.8	633	9	US-09-884-465A-363	Sequence 363, App
34	3287	61.8	633	9	US-09-884-465A-364	Sequence 364, App
35	3281	61.7	894	9	US-09-884-465A-336	Sequence 336, App
36	3278	61.7	633	9	US-09-884-465A-365	Sequence 365, App
37	3272	61.6	894	9	US-09-884-465A-340	Sequence 340, App
38	3272	61.6	894	9	US-09-884-465A-342	Sequence 342, App
39	3270.5	61.5	895	9	US-09-884-465A-345	Sequence 345, App
40	3269	61.5	633	9	US-09-884-465A-353	Sequence 353, App
41	3269	61.5	888	9	US-09-884-465A-371	Sequence 371, App
42	3268	61.5	889	9	US-09-884-465A-338	Sequence 338, App
43	3266.5	61.5	889	9	US-09-884-465A-348	Sequence 348, App
44	3251	61.2	627	9	US-09-884-465A-366	Sequence 366, App
45	3249	61.1	627	9	US-09-884-465A-367	Sequence 367, App

DB 201 YIVPHGHHYIPKSDLSASLAAKAGLAKNMQPSQLSYSTASDNNQTSVAKGSTSK 260
QY 241 PANKSENLSLKLKLYDSPAQRYSSDGLVDFPAKISRTNGVAIPHGDHYHFIPIYSK 300
DB 261 PANKSENLSLKLKLYDSPAQRYSSDGLVDFPAKISRTNGVAIPHGDHYHFIPIYSK 320
QY 301 LSALKEETARMPISGTSVSTNAKPNVSVSLGSLSNPSSLTTSKELSSASDGYIFN 360
DB 321 LSALKEETARMPISGTSVSTNAKPNVSVSLGSLSNPSSLTTSKELSSASDGYIFN 380
QY 361 PKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLPNNSLATPSPSLPINTPGTSHEKHEED 420
DB 381 PKDIVEETATAYIVRHGDHFIYIPKSNQIGOPTLPNNSLATPSPSLPINTPGTSHEKHEED 440
QY 421 GYGFDAIRIITAEDESGFVMSGHGDNHYFFKKDLTEBQIKAAQKHEEVKTSNGLDLSLSS 480
DB 441 GYGFDAIRIITAEDESGFVMSGHGDNHYFFKKDLTEBQIKAAQKHEEVKTSNGLDLSLSS 500
QY 481 HEQDYPGNKEMKOLDKRIEKGIMKQYGVKRESIVVKNKNAIYIPHGHHHADPTD 540
DB 501 HEQDYPGNKEMKOLDKRIEKGIMKQYGVKRESIVVKNKNAIYIPHGHHHADPTD 560
QY 541 EHKPVGIGHSHSNVLEFKPEGVAKKGNKRVYTGEEITNVVLLKNSFNQNFLLANGQ 600
DB 561 EHKPVGIGHSHSNVLEFKPEGVAKKGNKRVYTGEEITNVVLLKNSFNQNFLLANGQ 620
QY 601 KRVSFSPPELEKLGIMLVKLITPDGKLVKSVKGVCEGVGNIANFELDQVLPQGT 660
DB 621 KRVSFSPPELEKLGIMLVKLITPDGKLVKSVKGVCEGVGNIANFELDQVLPQGT 680
QY 661 FKYTIASKDPEVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAPVPGTDA 720
DB 681 FKYTIASKDPEVSYDGTFTVPTSLAYKMASQTIYFPFHAGDTYLRVNPQFAPVPGTDA 740
QY 721 VRVDEPHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 780
DB 741 VRVDEPHGNAYLENNYKVEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVE 800
QY 781 VPILKENQTDKPSILQFQRKNAQENSKLDEKVEPKPSEKVEKELSETNSNSTL 840
DB 801 VPILKENQTDKPSILQFQRKNAQENSKLDEKVEPKPSEKVEKELSETNSNSTL 860
QY 841 BEVPTVDPVQEVKAFKESYKMLERVLNMDGTIELYLPSEGEVKKNNMADFTGEAPOQN 900
DB 861 BEVPTVDPVQEVKAFKESYKMLERVLNMDGTIELYLPSEGEVKKNNMADFTGEAPOQN 920
QY 901 GENKPSNGKVSQTVENQNTENKPADSLPEAPNEKVPKPNSTDNGLNPNENGVSDPM 960
DB 921 GENKPSNGKVSQTVENQNTENKPADSLPEAPNEKVPKPNSTDNGLNPNENGVSDPM 980
QY 961 LDPALAEAPADVPQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVKKNLSDFA 1019
DB 981 LDPALAEAPADVPQEKLEKFTASYGLGLDSVIFNMDGTIELRPSGEVKKNLSDFA 1039

RESULT 2

US-09-884-465A-381
; Sequence 381, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 1238
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC.FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: xaa = Methionine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (430)..(430)
; OTHER INFORMATION: xaa = Glycine or nothing
; NAME/KEY: MISC.FEATURE
; LOCATION: (431)..(431)
; OTHER INFORMATION: xaa = proline or nothing
US-09-884-465A-381

Query Match 81.3%; Score 4322; DB 9; Length 1238;

Best Local Similarity 70.7%; Pred. No. 6.3e-244;

Matches 876; Conservative 39; Mismatches 64; Indels 260; Gaps 12;

QY 25 SOSQSKSEN--TPDOVSQKEGIAQBIQIKITDOGVYVTHGDNHYHNGKVPIDALFSE 82
DB 16 SQRHVESDGLIFDPAQITSR-----TARGVAVPHGNHYHF---IPVEQMSSEL 59
QY 83 EL-----LMKDPNPQLKDA-----DI 98
DB 60 EKRIARIPIRYSRNHWVDSRPEEPSQPTPEPSPQAPNPQAPNPIDEKLYKEA 119
QY 99 VNEVGKGIYIKVDG-KYYVLKD-----AAHADNVRTKDEI-----133
DB 120 VRKVGQGVYFEENGVSRYTPAKNLSAETAAGTDSKLAKQESLSHKLGAKTDLPSDREF 179
QY 134 -----NRQKQ-----EHVKD--NEKV-----147
DB 180 YNKAYDLLARIHQDLDNKGROVDFEALDNLLERLKDSSDKVLDVDDILAFLAPIRHP 239
QY 148 -----NSNVAVARSQRYTNGYVFNPAIDEDTGNAYIVPHGHHYHPIKS 195
DB 240 RLKQPNQIITYTODEIQVAKLAGKTYTDEGYIFDPRDITSDGDAYVTPHMTSHWKKD 299
QY 196 DLSASELAAAKAHLAGK-----212
DB 300 SLSEAPAAAQAVAKEKGLTPPSTPDHDSGNTAKGAEAIYNRVKAARKVPLDRMPYNLQ 359
QY 213 -----212
DB 360 YTVKNGSLIIPHYDHYHNKFEWDFEGLYEAPKGYTLLEDLATVKKYVHNPERHSD 419
QY 213 -----NMQPSQLSYSTASDNNQTSVAKGSTSKPANKSENLSLKLKLYDPS 260
DB 420 NGFGNASDHVXXNQPSQLSYSTASDNNQTSVAKGSTSKPANKSENLSLKLKLYDPS 479
QY 261 AQRYSESGLVDFPAKISRTNGVAIPHGDHYHFIPIYSKLSALEEKIARWVPISTGST 320
DB 480 AQRYSESGLVDFPAKISRTNGVAIPHGDHYHFIPIYSKLSALEEKIARWVPISTGST 539
QY 321 VSTNAKNEVSVSLGSLSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDH 380
DB 540 VSTNAKNEVSVSLGSLSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDH 599
QY 381 HYIPKSNQIGOPTLPNNSLATPSPSLPINTPGTSHEKHEEDGYGFDAIRIITAEDESGFVMS 440
DB 600 HYIPKSNQIGOPTLPNNSLATPSPSLPINTPGTSHEKHEEDGYGFDAIRIITAEDESGFVMS 659
QY 441 HGDHNYHFFKKDLTEBQIKAAQKHEEVKTSNGLDLSLSSHEQDYPGNKEMKOLDKRIE 500
DB 660 HGDHNYHFFKKDLTEBQIKAAQKHEEVKTSNGLDLSLSSHEQDYPGNKEMKOLDKRIE 719
QY 501 EKTAGINKQYGVKRESIVVKNKNAIYIPHGHHHADPTDEHKPVGIGHSHSNVLEFKPE 560
DB 561 EKTAGINKQYGVKRESIVVKNKNAIYIPHGHHHADPTDEHKPVGIGHSHSNVLEFKPE 599

Db 720 EKIAGIMKQYKVRKESIVYNNKNAIYPHGDHHDADIDEHKPVGIGHSHSNVLEFKPE 779
QY 561 EGVAKEGKNVYTGEEITNVVNLKSTNNQNTLANGOKRVSFPPPELEKKGGINML 620
Db 780 EGVAKEGKNVYTGEEITNVVNLKSTNNQNTLANGOKRVSFPPPELEKKGGINML 839
QY 621 VKLITPDGKLVKSVKVGEGVGNANPELDQPYLPQGTFTKTIASKDYPEVSDGTFT 680
Db 840 VKLITPDGKLVKSVKVGEGVGNANPELDQPYLPQGTFTKTIASKDYPEVSDGTFT 899
QY 681 VPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNKYG 740
Db 900 VPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEFGHGNAYLENNKYG 959
QY 741 EIKLPIPKLNOGTTAGNKPITVFEMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOK 800
Db 960 EIKLPIPKLNOGTTAGNKPITVFEMANAYLDNQSTYIIVEVPILEKENOTDKPSILPOK 1019
QY 801 RNKAQENSKLDEKVEPKTSEKVEKEKLGSETGNSTNSNLEEVPVDPVOEKVAFASY 860
Db 1020 RNKAQENSKLDEKVEPKTSEKVEKEKLGSETGNSTNSNLEEVPVDPVOEKVAFASY 1079
QY 861 GMLNVLNMGDTIELYLPSEGVIKKNMADFTGEAPQNGENKPSNGKSTGIVENOP 920
Db 1080 GMLNVLNMGDTIELYLPSEGVIKKNMADFTGEAPQNGENKPSNGKSTGIVENOP 1139
QY 921 TENKPADSLPEAPNEKPVKPNSTNGMLNPEGNYGSDPMLDPALEAPAVDPVQEKLEK 980
Db 1140 TENKPADSLPEAPNEKPVKPNSTNGMLNPEGNYGSDPMLDPALEAPAVDPVQEKLEK 1199
QY 981 FTASYGLGLDSVIFNMMDGTIELRLPSGEVVIKKNLSDFIA 1019
Db 1200 FTASYGLGLDSVIFNMMDGTIELRLPSGEVVIKKNLSDFIA 1238

RESULT 3

US-09-884-465A-10
; Sequence 10, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884.465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-10

Query Match 80.5%; Score 4276.5; DB 9; Length 840;
Best Local Similarity 82.3%; Pref. No. 1.7e-241;
Matches 835; Conservative 0; Mismatches 2; Indels 177; Gaps 1;

QY 1 CAYALNQHRSOENKDNRRVSVYDQSQQSKSENLPDQVSKQEGIQABQIVIKITDQGV 60
Db 1 CAYALNQHRSOENKDNRRVSVYDQSQQSKSENLPDQVSKQEGIQABQIVIKITDQGV 60
QY 61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQLKADIVNEVKGYIKVDGKYVYVILKD 120
Db 61 TSHGDHYHYNGKVPYDALFSEELLMDPNYQLKADIVNEVKGYIKVDGKYVYVILKD 120
QY 121 AAHADNVRTKDEINRQKOEHVKNVAVARSQGRYTTNDGYVFNPAIDITDIGNA 180

Db 121 AAHADNVRTKDEINRQKOEHVKNVAVARSQGRYTTNDGYVFNPAIDITDIGNA 180
QY 181 YIVPHGGHYHYTPKSDLSASELAARAKAHLAGKNMOPSOISYSSSTASDNNTOSVAKGTSK 240
Db 181 YIVPHGGHYHYTPKSDLSASELAARAKAHLAGKNMOPSOISYSSSTASDNNTOSVAKGTSK 240
QY 241 PANKSENLOSLKELYDPSAQRVSESDGLVDFPAKLIISRTPNGVAIPHGDHYHIFPYK 300
Db 224 ----- 223
QY 301 LSALEEKIARVPISTGTVSTVSTNAKNPNEVSSIGLSLSPSSLTTSKELSSASDGYIFN 360
Db 224 ----- 223
QY 361 PKDIVEETATAYIVRHGDHFHVPKSNIOIGOPTLPNNSLATPSPSLPINPGTSHKHEED 420
Db 224 -----TPSPSLPINPGTSHKHEED 243
QY 421 GYGFDPANRIIADESGFYMSHGDHNNHYEFFKDLTDEQIKAAQKHLKEEVKTSNHSIGDLSLS 480
Db 244 GYGFDPANRIIADESGFYMSHGDHNNHYEFFKDLTDEQIKAAQKHLKEEVKTSNHSIGDLSLS 303
QY 481 HEODYPGNAKEMKDLKIEEKIAGIKMKGKRESIVVNEKNAIYPHGDHHDADID 540
Db 304 HEQDTPSNAKEMKDLKIEEKIAGIKMKGKRESIVVNEKNAIYPHGDHHDADID 363
QY 541 EHKPVGIGHSHSNVLEFKPEEGVAKKEGKVTGEEITNVVNLKSTNNQNTLANGO 600
Db 364 EHKPVGIGHSHSNVLEFKPEEGVAKKEGKVTGEEITNVVNLKSTNNQNTLANGO 423
QY 601 KRVSFPPELEKKGGINMLVKLITPDGKLVKSVKVGEGVGNANFELDQPYLPQGT 660
Db 424 KRVSFPPELEKKGGINMLVKLITPDGKLVKSVKVGEGVGNANFELDQPYLPQGT 483
QY 661 FKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 720
Db 484 FKYTIASKDYPEVSDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDAL 543
QY 721 VRVDFEFGHGNAYLENNKYGKIEKLPKLNQGTTRTAGNKIPVTFEMANAYLDNQSTYIVE 780
Db 544 VRVDFEFGHGNAYLENNKYGKIEKLPKLNQGTTRTAGNKIPVTFEMANAYLDNQSTYIVE 603
QY 781 VPILEKENOTDKPSILPOKFKENKAQENSKLDEKVEEPEKTEKVEKEKLSGTGNSNSTL 840
Db 604 VPILEKENOTDKPSILPOKFKENKAQENSKLDEKVEEPEKTEKVEKEKLSGTGNSNSTL 663
QY 841 EEVPTVDPVOEKVAKFAESYGMKLENVLFNMMDGTIELYLPSEGVVIKKNMADFTGEAPQGN 900
Db 664 EEVPTVDPVOEKVAKFAESYGMKLENVLFNMMDGTIELYLPSEGVVIKKNMADFTGEAPQGN 723
QY 901 GENKPSNGKSTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGNYGSDPM 960
Db 724 GENKPSNGKSTGTVENOPTENKPADSLPEAPNEKPVKPNSTNGMLNPEGNYGSDPM 783
QY 961 LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMMDGTIELRLPSGEVVIKKNL 1014
Db 784 LDPALAEAPAVDPVQEKLEKFTASYGLGLDSVIFNMMDGTIELRLPSGEVVIKKNL 837

RESULT 4

US-09-884-465A-382
; Sequence 382, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044

Qy	25	SQSQSKENL--TPDOVSOKEGIQAEQIIVIKITDQGYVTSHGDDHYHNGKVPYDALFSE	82
Dz	16	SQRHVESDGLIFDPAGTISR-----TARGVAVPHGNHHF-----IPYQMSEL	59
Qy	83	EL-----LMKDPNYOLKDA-----DI	98
Dz	60	EKRIARIILRYSNHWPDSPREPSPOPTPEPSPQPAPNPQAPSNIIDEKLVAEA	119
Qy	99	VNEVGKYIIKVDG-KYYVLYKD-----AAHADNVTKDEI-----	133
Dz	120	VRKVGDGVFEENGVSRYIPAKNLSETAAGIDSKLAKOESISHLKGAKKTDLPPSSDREF	179
Qy	134	-----NRQKQ-----EHVKD--NEXY-----	147
Dz	180	YNKAYDLARIHQDLLDNKGROVDFEALDNLLERLDKDVSKVKLVDDLI AFLAPIRHPE	239
Qy	148	-----NSNAVARSQGRYTNDGYVFNPADIETGCNAVIYPGGHYHVPKS	195
Dz	240	RLQKPNAITYTDETOQVAKLAGKITTEDGYTFDPDITISDGDAYVTPTHMTHSHWIKKD	299
Qy	196	DLASASELAARAAAH-----	208
Dz	300	SLSEABRAAQAYAKEKGLTPPSTDHODSGNTEAKGAEEAIYNRVRAAKKVPDLRWPNLQ	359
Qy	209	-----	208
Dz	360	YTVEKNGSLIIPHVDYHNKIFEWTEGLYEAPKGYTLIEDLLATVKYVVEHPNERPHSD	419
Qy	209	-----	208
Dz	420	NGFCNASDHVORNKNQOADINTQEXPEEKPQTEKPEETPREEXPKQSEKPSKPTEEP	479
Qy	209	-----LACK-----	212
Dz	480	BEESPSESEEPQVETEKEVCKLEAREDDLJKIODPIKSNAKETTLGLKNLLFGTDQNN	539
Qy	213	-----NMQPSQLSYSTASDNNNTQSVAKGSTSKPANKSENLOSLLK	253
Dz	540	TIMAEKELLALLKESXXXNMWQSQLSYSTASDNNNTQSVAKGSTSKPANKSENLSLLK	599
Qy	254	ELYDSPSAQRYSESGLVFPDAKIIISRTPNGVAIPGHGDHYHFPTYSKLSALEEKIARMWP	313
Dz	600	ELYDSPSAQRYSESGLVFPDAKIIISRTPNGVAIPGHGDHYHFPTYSKLSALEEKIARMWP	659
Qy	314	ISGTGSTVSTYNAKENVVVSLGSLSNSPSSLTYSKELSSASDCGYFNPKDIVETATAYI	373

```

1  PUBLICATION NO. US20030077253A1
2
3  GENERAL INFORMATION:
4
5  APPLICANT: Shire Biochem, Inc.
6
7  APPLICANT: Hamel, Josee
8
9  APPLICANT: Brodeur, Bernard
10
11 APPLICANT: Martin, Denis
12
13 APPLICANT: Charland, Nathalie
14
15 APPLICANT: Ouellet, Catherine
16
17 TITLE OF INVENTION: Streptococcus Antigens
18
19 FILE REFERENCE: 055190-0044
20
21 CURRENT APPLICATION NUMBER: US/09/884,465A
22
23 CURRENT FILING DATE: 2001-06-20
24
25 PRIOR APPLICATION NUMBER: 60/212,683
26
27 PRIOR FILING DATE: 2000-06-20
28
29 NUMBER OF SEQ ID NOS: 384
30
31 SOFTWARE: PatentIn version 3.1
32
33 SEQ ID NO 378
34
35 LENGTH: 1378
36
37 TYPE: PRT
38
39 ORGANISM: Artificial Sequence
40
41 FEATURE:
42
43 OTHER INFORMATION: Unknown Organism
44
45 NAME/KEY: MISC_FEATURE

```



```

: LOCATION: {1}, {1}
: OTHER_INFORMATION: Xaa = Methionine or nothing
: NAME/KEY: MISC_FEATURE
: LOCATION: {570}..{570}
: OTHER_INFORMATION: Xaa = Glycine or nothing
: NAME/KEY: MISC_FEATURE
: LOCATION: {571}..{571}
: OTHER_INFORMATION: Xaa = Proline or nothing
US-09-884-465A-378

```

Query Match	80.1%	Score 4258.5;	DB 9;	Length 1378;
Best Local Similarity	63.7%	Pred. No. 3.7e-240;		
Matches 879;	Conservative 37;	Mismatches 61;	Indels 403;	Gaps 13;

QY	25	SQSQKSEN	--T	PDQVSQ	EGIOA	ELV	KITDQ	GVVTS	HG	DHYHY	NYK	VPYDAL	FS	82	
Db	17	SERHVESD	GLIFD	PAQ	TTSR	-----	--T	ARGVAV	PHG	NHYHF	-----	IBYEOMSEL	60		
QY	83	E	-----	-----	-----	-----	-----	LLMKDP	NYQ	-----	-----	LKDA	96		
Db	61	EKRTIARI	PLRYR	NHVV	PDSP	RPEQSP	OSTP	PSPS	LOP	APNP	QAPSP	NPIDEK	LVKEA	120	
QY	97	DIVNEV	GGYII	KVDG	-KY	VYI	LKD	-----	AAH	ADNV	RTKDEI	-----	133		
Db	121	--VRK	VDG	VYF	ENG	SVR	IPAK	DL	SAE	TAA	GIDSK	LAK	OE	178	
QY	134	-----	-----	-----	-----	-----	-----	NROK	-----	EHV	KD	-NEKV	-----	147	
Db	179	EFY	NKAYD	LLAR	I	HOD	LLDN	KQV	DE	VL	DNLLER	LK	QVSS	238	
QY	148	-----	-----	-----	-----	-----	-----	NSN	VAR	SQGR	YTTND	GV	FN	193	
Db	239	PERL	GK	NAQ	IT	T	DD	ETQ	IAK	LAK	GK	YTT	EDG	298	
QY	194	KSDL	SASE	LAA	AKA	HL	AGK	-----	-----	-----	-----	-----	-----	212	
Db	299	KDSL	SE	AE	RAA	AAQ	AYAK	EK	GL	TP	TDH	QD	SG	358	
QY	213	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	212	
Db	359	LQY	TVEY	K	NGS	L	I	PHY	D	HYH	N	I	K	418	
QY	213	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	212	
Db	419	SDNG	F	GN	ASD	HV	R	KN	AD	Q	S	K	PD	478	
QY	213	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	212	
Db	479	TDTE	ETEE	EA	ED	T	DE	AE	I	P	O	V	EN	538	
QY	213	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	212	
Db	539	LLG	T	KD	NN	T	I	AE	V	D	S	L	I	598	
QY	240	KP	AKSEN	L	Q	S	L	K	E	L	Y	D	S	PA	299
Db	599	KP	AKSEN	L	Q	S	L	K	E	L	Y	D	S	PA	658
QY	300	KLS	AE	PK	I	ARM	V	P	I	SG	T	SV	T	NAK	359
Db	659	KLS	AE	PK	I	ARM	V	P	I	SG	T	SV	T	NAK	718
QY	360	NP	KD	VE	ET	A	T	A	V	I	R	G	H	D	419
Db	719	NP	KD	VE	ET	A	T	A	V	I	R	G	H	D	778
QY	420	DG	Y	F	D	A	N	R	I	A	E	D	S	G	479
Db	779	DG	Y	F	D	A	N	R	I	A	E	D	S	G	838
QY	480	S	H	E	D	Y	P	G	N	A	K	E	M	D	539
Db	839	S	H	E	D	Y	P	G	N	A	K	E	M	D	898

QY 8 HRSQENKNNRVSYYVDSQSSQSKSENLTDPQVQSKEGIOAE----- 48
Db 67 HSHWIKKDSLSAERAAQAAYAKGKLTPTSDHODSGNTEAKGAEAIYNRVKAARKVPL 126
QY 49 -----QIVIKITDQGYVTSCHDHYHYNGKVPYDALPSFELLKMDPNYQLKADIVNE 101
Db 127 DRPMYNLOQTVVEKNGSLIPIHYDHYH--NIRPEW---FDEGLYAPKGYSLD----- 175
QY 102 VKGGYIIRKDGKYYV-YLKDAHAADN--VRTKDEINRQK--QEHVKONEKVNKNVAVARS 156
Db 176 -----LLATVKYVVEHPNERPHSDNGFCGNASHVRKKNKADQDQSKPDKEHDEV-----S 225
QY 157 QGRYTN-----GIVFNPA-----DIIDTGNAYIVPHGHHYHYPKSDL 197
Db 226 EPTHPSDEKENHAGLNPSADNLKYPSTDETEEAEDTTDEAEIPQ-----VENSVI 279
QY 198 SA-----SELAAR----- 206
Db 280 NAKIADAELKXVTPDSIFQNADETTLGLKSLLLGTGKDNNTISAEVDSLLALLKESQP 339
QY 207 AHLAGNMQPSQLSYSTASDNNQTSVAKGSTKPKANKSENLOSLKELYDPSAQRYSE 266
Db 340 APIQXXNMQPSQLSYSTASDNNQTSVAKGSTKPKANKSENLOSLKELYDPSAQRYSE 399
QY 267 SDGLVDPKALISRTPNGVAIPHGDHYHFTPYKLSALSALBEKATARMVPISGTGSTVSTNAK 326
Db 400 SDGLVDPKALISRTPNGVAIPHGDHYHFTPYKLSALSALBEKATARMVPISGTGSTVSTNAK 459
QY 327 PNEVSSLSGLSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKS 386
Db 460 PNEVSSLSGLSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFIYIPKS 519
QY 387 NQIQGQPTLPNNSLATPSPLIPNGTSHKHEBDGYPGDFANRIIAEDSGFVMSHGDHNN 446
Db 520 NQIQGQPTLPNNSLATPSPLIPNGTSHKHEBDGYPGDFANRIIAEDSGFVMSHGDHNN 579
QY 447 YFFKKDLTEEQIKAAQKHLBEVTSNGLDLSLSSHQDYPGNAKEMKDLKITEEKIAGI 506
Db 580 YFFKKDLTEEQIKAAQKHLBEVTSNGLDLSLSSHQDYPGNAKEMKDLKITEEKIAGI 639
QY 507 MKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEECVAKK 566
Db 640 MKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKPEECVAKK 699
QY 567 EGNKYVTGEELTNVNLKNTSTNNQNTFLANGKQKRVSTSFPPPELEKKGILNMLVKLITP 626
Db 700 EGNKYVTGEELTNVNLKNTSTNNQNTFLANGKQKRVSTSFPPPELEKKGILNMLVKLITP 759
QY 627 DGKVLKVSQKVGEGVGNANTANFELQYLPQGTFRYITASKDYPEVSDGTFTVPTSIA 686
Db 760 DGKVLKVSQKVGEGVGNANTANFELQYLPQGTFRYITASKDYPEVSDGTFTVPTSIA 819
QY 687 YKMASOTIFYPEHAGDTYLRVNPQFAYPKCTDALRVDFEFGHNALENKYGEIKLPI 746
Db 820 YKMASOTIFYPEHAGDTYLRVNPQFAYPKCTDALRVDFEFGHNALENKYGEIKLPI 879
QY 747 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSLIPQKRNKAQE 806
Db 880 PKLNOGTRTAGNKIPVTFMANAYLDNQSTYIIVEVPILEKENOTDKPSLIPQKRNKAQE 939
QY 807 NSKLDEKVEBPKESEKKEKLSGTGNTSTNLTLEEVPTVPQEVKAKFAESYGNKLEN 866
Db 940 NSKLDEKVEBPKESEKKEKLSGTGNTSTNLTLEEVPTVPQEVKAKFAESYGNKLEN 999
QY 867 VLFNMDGTIELYLPSSGVIKKNMADEFTGEAPQNGENKPKSNGKSVSTGTVENOPTENKPA 926
Db 1000 VLFNMDGTIELYLPSSGVIKKNMADEFTGEAPQNGENKPKSNGKSVSTGTVENOPTENKPA 1059
QY 927 DLSLEAPNEKXPVKENSTDNGLNPEGNVGSQDMLDPALAEAPVQOEKLEKFTASYG 986
Db 1060 DLSLEAPNEKXPVKENSTDNGLNPEGNVGSQDMLDPALAEAPVQOEKLEKFTASYG 1119

QY 987 LGLDSVIFNMDGTIELRLPSGEVVIKKNLSDFIA 1019
Db 1120 LGLDSVIFNMDGTIELRLPSGEVVIKKNLSDFIA 1152
RESULT 7
US-09-884-465A-370
; Sequence 370, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 370
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-370
Query Match 63.7%; Score 3384; DB 9; Length 906;
Best Local Similarity 72.7%; Pred. No. 2,1e-189;
Matches 691; Conservative 40; Mismatches 91; Indels 128; Gaps 12;
QY 148 NSNAVARSQRYITNDQYVFNPADIIETDGNAYIVPHGHHYHYPKSDLSASLAQA 207
Db 7 DDEIQVAKLAGRYTDEGYIF-----DTS-----WIKKDSLEAERAAQA 47
QY 208 HLAGNMQPSQLSYSTASDNNQTSVAKGSTKPKANKSENLOSLKELYDPSAQRYSES 267
Db 48 YAKERGLIP-----PSTDHODSGNTEAKGA-----EALYNRVKAARKVPL 87
QY 268 DGLVDPKALISRTPNGVAIPHGDHYHFIYIPYS-----KLSALBEKATARM--VPI 314
Db 88 DRMPYNLOYTVEVKNGSLIIPHYDHYHNIKFEWDEGLYAPKGYSLDGLATVYIYEP 147
QY 315 SGTGSTVSTN-----AKPNEV-----SSLSGLSN-----PSSITTS 347
Db 148 RNASHVRKNKADQSKPDEKHEHDESEPTHPESDEKENHAGLNPSADNLKYPSTDEE 207
QY 348 KE-----LSSASDGYIFNPKD-----IVETA 369
Db 208 TEEAEADTTDEAEIPGTPSIQNADETTLGLKSLLLGTGKDNNTISAEVDSLLALLKESQ 267
QY 370 TAYIVRHGDHFIYIPKSNOIGOPTLPNNSLATPSPLIPNGTSHKHEEDGYGDFANRI 429
Db 268 PAPI-----QGQIQGQPTLPNNSLATPSPLIPNGTSHKHEEDGYGDFANRI 316
QY 430 IAEDESGFVMSHGDHNNHFFKKDLTEEQIKAAQKHLBEVTSNGLDLSLSSHQDYPGNA 489
Db 317 IAEDESGFVMSHGDHNNHFFKKDLTEEQIKAAQKHLBEVTSNGLDLSLSSHQDYPGNA 376
QY 490 KEMKDLKITEEKIAGIMKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGH 549
Db 377 KEMKDLKITEEKIAGIMKQYGVKRESIVNKEKNALIIYPHGDHHDADPIDEHKPVGIGH 436
QY 550 SHSNYELFKPEECVAKKKEGKNKYVTGEELTNVNLKNTSTNNQNTFLANGKQKRVSTFPP 609
Db 437 SHSNYELFKPEECVAKKKEGKNKYVTGEELTNVNLKNTSTNNQNTFLANGKQKRVSTFPP 496
QY 610 ELEKKGILNMLVKLITPDKVLEKVSQKVGEGVGNANTANFELQYLPQGTFRYITASKD 669

```
Db 497 ELEKLGINMLVKLITPDGKVLKGVKGVGEGVGNIANFELDPQYLPGGTFFKTTASKD 556
QY 670 YPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHG 729
Db 557 YPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHG 616
QY 730 NAYLENNYKVGELKIPKLPKLNQCTTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQ 789
Db 617 NAYLENNYKVGELKIPKLPKLNQCTTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQ 676
QY 790 TDKPSILPQPKRKAQENSKLDEKVEEPTSEKVEKEKLSSETNSTLEEVPTVDPV 849
Db 677 TDKPSILPQPKRKAQENSKLDEKVEEPTSEKVEKEKLSSETNSTLEEVPTVDPV 736
QY 850 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVVKKNLSDFIA 1019
Db 737 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVVKKNLSDFIA 906
QY 910 KVESTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVGSDDPMLDPALEAP 969
Db 797 KVESTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVGSDDPMLDPALEAP 856
QY 970 AVDPVQEKLEKFTASYGLGLDLSVIFNMDGTIELRPSGEVVKKNLSDFIA 1019
Db 857 AVDPVQEKLEKFTASYGLGLDLSVIFNMDGTIELRPSGEVVKKNLSDFIA 906

RESULT 8
US-09-884-465A-332
; Sequence 332, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Jossee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 332
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-332

Query Match 63.5%; Score 3375; DB 9; Length 906;
Best Local Similarity 72.6%; Pred No. 7e-189;
Matches 690; Conservative 40; Mismatches 92; Indels 128; Gaps 12;

QY 148 NGNAVARSQGRYTTNDGVYFNPADIETGNAYIVPHSGHYHYIPKSDLSASELAATAA 207
Db 7 DDEIOVAKLAGKYTTEDGYIF-----DTS-----WIKKDSLSAEARAAQA 47
QY 208 HLAGKNMQPSQLSYSTASDNNQTSQVAKSTSKPANKSENLOSLLKELYDPSAQRYSES 267
Db 48 YAKEKGLTP-----PSTDHQDSGNTFEAKG-----EATYNRYKAAKRYPL 87
QY 268 DGLVFDPAKIIKRTNPNVAIPHCDHYHPTYS-----KLSALEKIAARM---VPI 314
Db 88 DRMPYNLQYTVKNGSLIIPSYDHYHNLKFWFDEGLYEPKGYSLDGLATVAKYIYEP 147
QY 315 SGTGTSTVSTN-----AKNEVY-----SSGLSSN---PSSLTTS 347
```

```
Db 148 RNASDHVRKANKADQDSKDEKDEKHEDEVSEPTHPSEDEKENHAGLNPSADNLYKPTSTDEE 207
QY 348 KE-----LSSASDGVIFPKD-----TWEEETA 369
Db 208 TEEEAEDTTDEAEIETPSIRONAMETITGLKSSLLLTGKDNNTISAEVDSLLALLKESQ 267
QY 370 TAYIVRHGDHPIHYIPKSNQIQOFTLPNNSLATPSPSLPINPGTSHKHEEDGYGFANDRI 429
Db 268 PAPI-----QGPQIGQFTLPNNSLATPSPSLPINPGTSHKHEEDGYGFANDRI 316
QY 430 IADESGFVMSHGDHNHYFFKKDLTEEOIKAAQKLEBKVKTSHNGLDLSLSHEQDYPGNA 489
Db 317 IADESGFVMSHGDHNHYFFKKDLTEEOIKAAQKLEBKVKTSHNGLDLSLSHEQDYPGNA 376
QY 490 KEMKOLDKIEBKIAIGIMKQYGVKRESIVVKNKXAIYPHGDHHDADPIDEHKPVGIGH 549
Db 377 KEMKOLDKIEBKIAIGIMKQYGVKRESIVVKNKXAIYPHGDHHDADPIDEHKPVGIGH 436
QY 550 SHSNYELKPEEGVAKKGNKVITGEELTNVNNLLKNSTFNQNTLANGOKRVYSFSEPP 609
Db 437 SHSNYELKPEEGVAKKGNKVITGEELTNVNNLLKNSTFNQNTLANGOKRVYSFSEPP 496
QY 610 ELEKLGINMLVKLITPDGKVLKGVGEGVGNIANFELDPQYLPGGTFFKTTASKD 669
Db 497 ELEKLGINMLVKLITPDGKVLKGVGEGVGNIANFELDPQYLPGGTFFKTTASKD 556
QY 670 YPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHG 729
Db 557 YPEVSYDGTFTVPTSLAYKMASQTIFFPHAGDTYLRVNPQFAVPKGTDALVRVDFEHG 616
QY 730 NAYLENNYKVGELKIPKLPKLNQCTTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQ 789
Db 617 NAYLENNYKVGELKIPKLPKLNQCTTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQ 676
QY 790 TDKPSILPQPKRKAQENSKLDEKVEEPTSEKVEKEKLSSETNSTLEEVPTVDPV 849
Db 677 TDKPSILPQPKRKAQENSKLDEKVEEPTSEKVEKEKLSSETNSTLEEVPTVDPV 736
QY 850 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVVKKNLSDFIA 1019
Db 737 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVVKKNLSDFIA 906
QY 910 KVESTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVGSDDPMLDPALEAP 969
Db 797 KVESTGTVENQPTENKPADSLPEAPNEKPKVPENSTDNGLNPEGNVGSDDPMLDPALEAP 856
QY 970 AVDPVQEKLEKFTASYGLGLDLSVIFNMDGTIELRPSGEVVKKNLSDFIA 1019
Db 857 AVDPVQEKLEKFTASYGLGLDLSVIFNMDGTIELRPSGEVVKKNLSDFIA 906

RESULT 9
US-09-884-465A-372
; Sequence 372, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Jossee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 372
; LENGTH: 906
; TYPE: PRT
```

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
US-09-884-465A-372

Query Match 63.3%; Score 3365; DB 9; Length 906;

Best Local Similarity 72.5%; Pred. No. 2.7e-188;

Matches 689; Conservative 40; Mismatches 93; Indels 128; Gaps 12;

```
QY 148 NSNVAVARSGRYTTNDGYVFNPAIDIEDTGNAIIVPHGGHYHYIPKSDLSASELAATAKA 207
DB 7 DDEIQVAKLAGKTYTDEGYIF-----DTS-----WIKKDSLSAEERAAQA 47
QY 208 HLAKKNQPSOLSYSSTASDNNTQSVAKGSTKPAKSENQSLKELLYDPSAQRYSIS 267
DB 48 YAKEKGLTP-----PSTDHDSGNGTEAKGA-----EAIYNRVKAARKVPL 87
QY 268 DGLVDFPAKIISRTPNGVAIPHGHGHIYFIPYS-----KLSALEEKIARM--VPI 314
DB 88 DEMPYNLOYTVEVKGSLIIPHYDHYENIKFEWFDEGLYEAPKGYSLDGLLATVRYVVEP 147
QY 315 SGTGSTVSTN-----AKPNEVV-----SSLGSLSSN-----PSLITS 347
DB 148 RNASHVRKKNKADQSKPDEDKEHDEVESEPTHEPSEDEKENHAGLNPSADNLYKPTDTEE 207
QY 348 KE-----LSSASDGYIFNPKD-----IVEETA 369
DB 208 TEEEAEDTTDEAEIPGTPSIRONAMETLTGLKSSLLGLTKDNNTISAEVDSLLALLKESQ 267
QY 370 TAYIVRHGDHFIYIPKSNQIQPTLPNNLSLTPSLPINTPGTSHKHEEDGYGFDANRI 429
DB 268 PAPI-----QGPQIQGPTLPNLSLTPSLPINTPGTSHKHEEDGYGFDANRI 316
QY 430 IAEDESFGVMSHGDNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSHEEQDYPGNA 489
DB 317 IAEDESFGVMSGSDSNHYFFKKDLTEQIKAAQKHLEEVKTSNGLDLSLSHEEQDYPGNA 376
QY 490 KEMKDLKKIEEKTAGIMQYGVKRESIVVNEKNAIIVPHGDHHDADPIDEKHPVGIQH 549
DB 377 KEMKDLKKIEEKTAGIMQYGVKRESIVVNEKNAIIVPHGDHHDADPIDEKHPVGIQH 436
QY 550 SHSNYELFKPEEGVAKREGNRYTGESLTIVNLLKNSTFNNGNFTLANGQKRVSFSEPP 609
DB 437 SHSNYELFKPEEGVAKREGNRYTGESLTIVNLLKNSTFNNGNFTLANGQKRVSFSEPP 496
QY 610 ELEKKGINMLVLIITPDGKLEKVSQKVGEGVGNANFELDQYLPQGTQKTYTIAKD 659
DB 497 ELEKKGINMLVLIITPDGKLEKVSQKVGEGVGNANFELDQYLPQGTQKTYTIAKD 556
QY 670 YPEVSYDGTFTVPTSLAYKMASQTIYPPFHAGDTYLVRNPQFAPVPGTDLVRVDFEHRG 729
DB 557 YPEVSYDGTFTVPTSLAYKMASQTIYPPFHAGDTYLVRNPQFAPVPGTDLVRVDFEHRG 616
QY 730 NAYLENNYKVEIKLPTKLNQGTTRTAGNKIIVPTWANAAYLDNQSYIIVEPILKEHQ 789
DB 617 NAYLENNYKVEIKLPTKLNQGTTRTAGNKIIVPTWANAAYLDNQSYIIVEPILKEHQ 676
QY 790 TDKPSILPQFRRKNAQENSKLDEKVEEPKTSKEVKEKLSFETGNSTNSNLTLEEVPTDVP 849
DB 677 TDKPSILPQFRRKNAQENSKLDEKVEEPKTSKEVKEKLSFETGNSTNSNLTLEEVPTDVP 736
QY 850 QEKVAKFAESYGMKLENVFNMDGTIELYLPSCGEVTKKNMADFTGEAPONGENKPSENG 909
DB 737 QEKVAKFAESYGMKLENVFNMDGTIELYLPSCGEVTKKNMADFTGEAPONGENKPSENG 796
QY 910 KVSQGTVENQPTENKPADSLPEAPNEKPVKPSNSTONGMLNPNCGNYSQDPMPLDPALEAP 969
DB 797 KVSQGTVENQPTENKPADSLPEAPNEKPVKPSNSTONGMLNPNCGNYSQDPMPLDPALEAP 856
QY 970 AVDPVQEKLEKFTASYGLGLDYSIFNMDGTIELRLPSGEVTKKNLSDFTA 1019
DB 857 AVDPVQEKLEKFTASYGLGLDYSIFNMDGTIELRLPSGEVTKKNLSDFTA 906
```

RESULT 10

US-09-884-465A-334
Sequence 334, Application US/09884465A
Publication No. US20030077293A1

GENERAL INFORMATION:

APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Chuellet, Nathalie
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 334
LENGTH: 900

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Unknown Organism

US-09-884-465A-334

Query Match 63.3%; Score 3362; DB 9; Length 900;

Best Local Similarity 74.7%; Pred. No. 4e-188;

Matches 699; Conservative 35; Mismatches 105; Indels 96; Gaps 15;

```
QY 148 NSNVAVARSGRYTTNDGYVFNPAIDIEDTGNAIIVPHGGHYHYIPKSDLSASELAATAKA 207
DB 7 DDEIQVAKLAGKTYTDEGYIF-----DTS-----WIKKDSLSAEERAAQA 47
QY 208 HLAKKNQPSOLSYSSTASDNNTQSVAKGSTS-----KPANK-----SENLO----- 249
DB 48 YAKEKGLTP-----PSTDHDSGNGTEAKGAELIYNVKAARKVPLDRMPYNLQYTVYKVN 102
QY 250 -SLI-----KELYDPSAQRYSSESGLV-----PDPAKIISRTPNGVA-----TPH 289
DB 103 GSLLIIPNIKFEWDEGLYEAPKG--YSLEDLATVRYVPEPRNASCHVRKKNKADQSKPD 160
QY 290 GDHVVH--FIPYKSLSALEEKIARMVPISTGTSTVSTNAKPNEV-----VSSLGSLSSNPS 342
DB 161 EDKHDVSEPTHEPSEDEKENHAGLNPSADNLYKPTDTEETEEAEQTTDRAEIPGTPS 220
QY 343 SLTTSKE--LSSASDGYIFNPKD-----IVEETATAYIVRHGDHFIYIPKSNQ 388
DB 221 IRQAMETLTGLKSSLLGLTKDNNTISAEVDSLLALLKESQAPI-----QQPQ 269
QY 389 IGQPTLPNNLSLTPSLPINTPGTSHKHEEDGYGFDANRIIAEDESFGVMSHGDNHYF 448
DB 270 IGQPTLPNNLSLTPSLPINTPGTSHKHEEDGYGFDANRIIAEDESFGVMSHGDNHYF 329
QY 449 FKKDLTEQIKAAQKHLEEVKTSNGLDLSLSHEEQDYPGNAKEMKDLKKIEEKTAGIMK 508
DB 330 FKKDLTEQIKAAQKHLEEVKTSNGLDLSLSHEEQDYPGNAKEMKDLKKIEEKTAGIMK 389
QY 509 QYGVKRESIVVNEKNAIIVPHGDHHDADPIDEHKVGVGIHSHSNYELFKPEEGVAKREG 568
DB 390 QYGVKRESIVVNEKNAIIVPHGDHHDADPIDEHKVGVGIHSHSNYELFKPEEGVAKREG 449
QY 569 NKVYTGELNVNVLKNSFTFNQNETLANGQKRVSFSPPELEKKGINMLVLIITPDG 628
DB 450 NKVYTGELNVNVLKNSFTFNQNETLANGQKRVSFSPPELEKKGINMLVLIITPDG 509
QY 629 KVLKVSQKVGEGVGNANFELDQYLPQGTQKTYTIAKSKDYPEVSYDGTFTVPTSLAYK 688
DB 510 KVLKVSQKVGEGVGNANFELDQYLPQGTQKTYTIAKSKDYPEVSYDGTFTVPTSLAYK 569
QY 689 MASQTIYPPFHAGDTYLVRNPQFAPVPGTDLVRVDFEHRGNAYLENNYKVEIKLPTPK 748
```

```
Db 570 MASQIFYPFHAGDTYLRVNPQFAVPKGTDALRVDFDEHGNAYLENNYKVGGEIKLPIPK 629
Qy 749 LNOGTRTAGNKIPVTFMANAYLDNOSTYIVVEPYLEKENQTDKPSILPOFKRNKAQENS 808
Db 630 LNOGTRTAGNKIPVTFMANAYLDNOSTYIVVEPYLEKENQTDKPSILPOFKRNKAQENS 689
Qy 809 KLDKVEEPTSEKVEKEKLSSTGNSNSTLEEYPTVDPVQEKVAKFAESYGMKLENVL 868
Db 690 KLDKVEEPTSEKVEKEKLSSTGNSNSTLEEYPTVDPVQEKVAKFAESYGMKLENVL 749
Qy 869 FNMDSGTIELYPSGEVKKNMADFTGEAPQNGENKPSGNGKSVSTGVENQNTENKPADS 928
Db 750 FNMDSGTIELYPSGEVKKNMADFTGEAPQNGENKPSGNGKSVSTGVENQNTENKPADS 809
Qy 929 LPEAPNEKVPKPSNSTDNGMLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLG 988
Db 810 LPEAPNEKVPKPSNSTDNGMLNPEGNVSDPMLDPALEAPAVDPVQEKLEKFTASYGLG 869
Qy 989 LDSVIFNMDGTIELRLPSGEVKKNLSDFIA 1019
Db 870 LDSVIFNMDGTIELRLPSGEVKKNLSDFIA 900

RESULT 11
US-09-884-465A-343
; Sequence 343, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 343
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-343

Query Match 62.9%; Score 3345.5; DB 9; Length 901;
Best Local Similarity 73.1%; Pred. No. 3.6e-187;
Matches 687; Conservative 37; Mismatches 103; Indels 113; Gaps 12;

Qy 148 NSNVAVSQGYTNDGKVFNPADIIEDTGNAVIVPHGGHYHYTPKSDLSASELAARA 207
Db 7 DDEIQVAKIAGRTTDDGYIF-----DTS-----WIKDLSLSEAAAAQA 47
Qy 208 HLAGNMPQSLSYSTASDNNQTSVAKGSTKSPANKSENLSLKLKELYDPSAQRYSSES 267
Db 48 YAKEGLIP-----PSTDHQDSGNTAKGA-----EAYNRVKAARKYPL 87
Qy 268 DGLVDPDAKISRTPNGVAIPHGDHYHIFPYS-----KLSALEKTIARM---VPI 314
Db 88 DRMPYNLQYTVKNGSLIIPHIDYHNIKFWFDEGLYEAAPKGYSLDILLATVAYVYEP 147
Qy 315 STGSTVSTN-----AKNEVV-----SSLSLSN-----PSSLTS 347
Db 148 RNASDHVRKNKADQSKPDEKDEHDEVSEPTHPSEDEKENHAGNPSADNLYKPSSTDTEE 207
Qy 348 KE-----LSSASDGYIFNPKD-----LVEETATAYIVRHGDH 379
```

```
Db 208 TEEAEDTTDAEIPGTPTSIQNAWETITGLKSSLLLTGKDNNTISAEDVSLALLKESQ 267
Qy 380 PHYIPKSNQIQPTLPNNSLATPSPSLPINFSTSHKHEEDGYGFDANRIIAEDSEGFVM 439
Db 268 PAPTQSGDQIQPTLPNNSLATPSPSLPINFSTSHKHEEDGYGFDANRIIAEDSEGFVM 327
Qy 440 SHGDHNYFFKDKDTEEQIKAAQKHLEEVKTSHTNGDLSLSHSEQDYPGNAKEMKDLKKI 499
Db 328 S-----YFFKDKDTEEQIKAAQKHLEEVKTSHTNGDLSLSHSEQDYPGNAKEMKDLKKI 381
Qy 500 BEKTAGIMKQGVKRESIVVNKEKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKP 559
Db 382 BEKTAGIMKQGVKRESIVVNKEKNATIIYPHGDHHDADPIDEHKPVGIGHSHSNYELFKP 441
Qy 560 BEGVAKEGKNVYTGHELTINYNLLKNSTFNNQNTLANGQKRVSFSPFPPELEKKGLTNM 619
Db 442 BEGVAKEGKNVYTGHELTINYNLLKNSTFNNQNTLANGQKRVSFSPFPPELEKKGLTNM 501
Qy 620 LVKLITPDGKYLEKSVGKVFGEVGNITANFELDQYLPFGQTFKTYITASKDYPEVSYDGT 679
Db 502 LVKLITPDGKYLEKSVGKVFGEVGNITANFELDQYLPFGQTFKTYITASKDYPEVSYDGT 561
Qy 680 TVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALRVDFDEHGNAYLENNYK 739
Db 562 TVPTSLAYKMASQTIFFYPFHAGDTYLRVNPQFAVPKGTDALRVDFDEHGNAYLENNYK 621
Qy 740 GEIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNOSTYIVVEPYLEKENQTDKPSILPQF 799
Db 622 GEIKLPIPKLNOGTRTAGNKIPVTFMANAYLDNOSTYIVVEPYLEKENQTDKPSILPQF 681
Qy 800 KRKAQENSKLDEKVEEPTSEKVEKEKLSSTGNSNSTLEEYPTVDPVQEKVAKFAES 859
Db 682 KRKAQENSKLDEKVEEPTSEKVEKEKLSSTGNSNSTLEEYPTVDPVQEKVAKFAES 741
Qy 860 YGMKLENVLFNMDGTIELYPSGEVKKNMADFTGEAPQNGENKPSGNGKSVSTGVENQ 919
Db 742 YGMKLENVLFNMDGTIELYPSGEVKKNMADFTGEAPQNGENKPSGNGKSVSTGVENQ 801
Qy 920 PTENKPADSLPEAPNEKVPKPSNSTDNGMLNPEGNVSDPMLDPALEAPAVDPVQEKLE 979
Db 802 PTENKPADSLPEAPNEKVPKPSNSTDNGMLNPEGNVSDPMLDPALEAPAVDPVQEKLE 861
Qy 980 KFTASYGLGDSVIFNMDGTIELRLPSGEVKKNLSDFIA 1019
Db 862 KFTASYGLGDSVIFNMDGTIELRLPSGEVKKNLSDFIA 901

RESULT 12
US-09-884-465A-335
; Sequence 335, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 335
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-335
```

[illegible]

```
Db 611 NAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTVIVEVPILEKENQ 670
QY 790 TDKPSILPQFKKNAQENSKLDEKVEEPTSEKVEKEKLSGTNSTSTSLTEEVPTDVPV 849
Db 671 TDKPSILPQFKKNAQENSKLDEKVEEPTSEKVEKEKLSGTNSTSTSLTEEVPTDVPV 730
QY 850 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNVADFTGEAPQNGENKPSNG 909
Db 731 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNVADFTGEAPQNGENKPSNG 790
QY 910 KYSTGTVENOPTENKPADSLPEAPNEKPKVYPENSTDNMGLNPEGVGSDPMDPALEAP 969
Db 791 KYSTGTVENOPTENKPADSLPEAPNEKPKVYPENSTDNMGLNPEGVGSDPMDPALEAP 850
QY 970 AVDPVQEKLEKFTASGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1019
Db 851 AVDPVQEKLEKFTASGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 900

RESULT 14
US-09-884-465A-341
: Sequence 341, Application US/09884465A
: Publication No. US2003007293A1
: GENERAL INFORMATION:
: APPLICANT: Shire Biochem, Inc.
: APPLICANT: Brodeur, Bernard
: APPLICANT: Martin, Denis
: APPLICANT: Charland, Nathalie
: APPLICANT: Ouellet, Catherine
: TITLE OF INVENTION: Streptococcus Antigens
: FILE REFERENCE: 055190-0044
: CURRENT APPLICATION NUMBER: US/09/884,465A
: PRIOR FILING DATE: 2001-06-20
: PRIOR APPLICATION NUMBER: 60/212,683
: NUMBER OF SEQ ID NOS: 384
: SOFTWARE: Patent in version 3.1
: SEQ ID NO 341
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Unknown Organism
US-09-884-465A-341

Query Match 62.6%; Score 3329; DB 9; Length 900;
Best Local Similarity 72.1%; Pred. No. 3.3e-186;
Matches 685; Conservative 40; Mismatches 91; Indels 134; Gaps 13;

QY 148 NSNVAVSQGYTTNDGCVFNPADIIEDTGNAYIVPHGGHYHYPKSDLSASELAARAKA 207
Db 7 DDEIQVAKLAGRYTTEDGYIF-----DTS-----WIKKDSLSEAEERAAQA 47
QY 208 HLAGKNMOPSQLSSTASDNSTQSVAKGSKPANKSENLSLKLKELYDSPSAQRYSES 267
Db 48 YAKEGLTP-----PSTDHQDSGNETAKA-----EAYNVRKAAKKVPL 87
QY 268 DGLVDPAKIIERTNGVAIPHGDHYHYPYS-----KLSAEKTIARM---VPI 314
Db 88 DRMPYNLOQTVKNGSLIIPSYDHYSNIKFEWDEGLYAPKGYSLDGLLATKYVYEP 147
QY 315 SGTGSTVSTN-----AKPNEVV-----SSLGSLSN---PSSLITS 347
Db 148 RNASHVRKKNADQSKPDEKDEKHEVSEPTHPESDEKENHAGLNPSADNLYKPKSTDTEE 207
QY 348 KE-----LSSASDGYIFNPKD-----IVEETA 369
Db 208 TEEEAEDTTDEAEIPCTPSIRONAMETLGLKSSULLGTCKDNNTISAEVDSLLALLKESQ 267
QY 370 TAYIVRHGDHFHYIPKSNQIGOPTLPNNSLATPSPSLINPCTSHKEHEEDGYGFEDANRI 429
```

```
Db 268 PAPI-----QGQIQOQTLPPNNSLATPSPSLINPCTSHKEHEEDGYGFEDANRI 316
QY 430 IAEDESSEVMGSHDHNHYFFKKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYFCNA 489
Db 317 IAEDESSEFVMS-----YFEKKDLTEEQIKAAQKHLEEVKTSNGLDLSLSSHEQDYFCNA 370
QY 490 KEMDLDKKIEEKTAGIMKQYGVKRESIVVANKENALIIYPHGDHHDADPTDEHKPVGIGH 549
Db 371 KEMDLDKKIEEKTAGIMKQYGVKRESIVVANKENALIIYPHGDHHDADPTDEHKPVGIGH 430
QY 550 SHSNYELFKPEEGVAKKGNKVYTGEEELTNVNLKKNSTENNQNFTILANGKRVSEFSEFP 609
Db 431 SHSNYELFKPEEGVAKKGNKVYTGEEELTNVNLKKNSTENNQNFTILANGKRVSEFSEFP 490
QY 610 ELEKKGILNMLVKLITPDGKVLKESKGVFGEVGNANTANFELDQPYLPQGTFKTYIASKD 669
Db 491 ELEKKGILNMLVKLITPDGKVLKESKGVFGEVGNANTANFELDQPYLPQGTFKTYIASKD 550
QY 670 YPEVSYDGTFTVPTSLAYKMASQITFIYFFHAGDTYLRVNPQFAVPKGTDALVRVDFEHG 729
Db 551 YPEVSYDGTFTVPTSLAYKMASQITFIYFFHAGDTYLRVNPQFAVPKGTDALVRVDFEHG 610
QY 730 NAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTVIVEVPILEKENQ 789
Db 611 NAYLENNYKVGELKIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTVIVEVPILEKENQ 670
QY 790 TDKPSILPQFKKNAQENSKLDEKVEEPTSEKVEKEKLSGTNSTSTSLTEEVPTDVPV 849
Db 671 TDKPSILPQFKKNAQENSKLDEKVEEPTSEKVEKEKLSGTNSTSTSLTEEVPTDVPV 730
QY 850 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNVADFTGEAPQNGENKPSNG 909
Db 731 QEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKNVADFTGEAPQNGENKPSNG 790
QY 910 KYSTGTVENOPTENKPADSLPEAPNEKPKVYPENSTDNMGLNPEGVGSDPMDPALEAP 969
Db 791 KYSTGTVENOPTENKPADSLPEAPNEKPKVYPENSTDNMGLNPEGVGSDPMDPALEAP 850
QY 970 AVDPVQEKLEKFTASGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 1019
Db 851 AVDPVQEKLEKFTASGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDFIA 900
```

RESULT 15

US-09-884-465A-345
: Sequence 345, Application US/09884465A
: Publication No. US2003007293A1
: GENERAL INFORMATION:

: APPLICANT: Shire Biochem, Inc.
: APPLICANT: Hamel, Josee
: APPLICANT: Brodeur, Bernard
: APPLICANT: Martin, Denis
: APPLICANT: Charland, Nathalie
: APPLICANT: Ouellet, Catherine
: TITLE OF INVENTION: Streptococcus Antigens
: FILE REFERENCE: 055190-0044
: CURRENT APPLICATION NUMBER: US/09/884,465A
: CURRENT FILING DATE: 2001-06-20
: PRIOR FILING DATE: 2001-06-20
: PRIOR APPLICATION NUMBER: 60/212,683
: NUMBER OF SEQ ID NOS: 384
: SOFTWARE: Patent in version 3.1
: SEQ ID NO 345
: LENGTH: 901
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Unknown Organism
US-09-884-465A-345

Query Match 62.6%; Score 3327.5; DB 9; Length 901;
Best Local Similarity 72.9%; Pred. No. 4.1e-186;
Matches 685; Conservative 37; Mismatches 105; Indels 113; Gaps 12;

Tue May 13 14:29:50 2003

```

QY 148 NSNVAARSQRYTNDGYVFNPAIIEDTGNAYIVPHGHHYVIPKSDLSASELAAXA 207
DB 7 DDEIQVAKLAGKYTTEDGYIF-----DTS-----WIKKDSLSEAEARAAQA 47
QY 208 HLAGKNMPSQLSYSTASDNSTQSVAGSKPANKSENLOSLLKELYDSPSAQRYSSES 267
DB 48 YAKERGLTP-----PSTDQDSGNTAEKGA-----EAIYNRVKAAKKVPL 87
QY 268 DGLVDPKAKIISRTTNGVAIPHGDVHPIVPS-----KLSALEEKIARM---VPI 314
DB 88 DRMPYNLQTVVEKNGSLIPIYSOSYHNKFEWFEDEGLYEPAPKGYSLLEDLATVKKYVEP 147
QY 315 SGTGSTVTSTN-----AKPNEVY-----SSLGSLSSN---PSSLTTS 347
DB 148 RNASDHVRKNKADQSKDPEDDEKHEDEVSEPTHPESDEKENHAGLNPSADNLYKPSTDTEE 207
QY 348 KE-----LSSASDGYIFNPKO---IVEETATAYIVRHGDH 379
DB 208 TEEAEADTTDEAEIPGTPSIRONAMETUTGLKSSLLTCKONNTISAEVDSLALLKESQ 267
QY 380 FHYIPKSNIGOPTLPNNSLATPSPSLPINPCTSHEKHEEDGYGFDANRIIAEDES GFVM 439
DB 268 PAPIQSGPOIGOPTLPNNSLATPSPSLPINPCTSHEKHEEDGYGFDANRIIAEDES GFVM 327
QY 440 SHGDHNYFFKKDLTEEQIKAAQKHLEEVKTSHNGLDLSLSHEQDYPGNAKEMKOLDKKI 499
DB 328 S-----YFFKKDLTEEQIKAAQKHLEEVKTSHNGLDLSLSHEQDYPGNAKEMKOLDKKI 381
QY 500 EEKIAGIMKQYGVKRESIVVNKEKNALIIYPHGDHHAADPIDEHKPVGICHSHSNYELFKP 559
DB 382 EEKIAGIMKQYGVKRESIVVNKEKNALIIYPHGDHHAADPIDEHKPVGICHSHSNYELFKP 441
QY 560 EECVAKKEGKNVYTGELTNVNVNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINM 619
DB 442 EECVAKKEGKNVYTGELTNVNVNLLKNSTFNQNTFLANGQKRVSFSPPELEKLGINM 501
QY 620 LVXLIITPDGKVLKESVKGVEGVGNIANFELDQPYLPQOTFKYTIASKDYPEVSYDGTFF 679
DB 502 LVKLIITPDGKVLKESVKGVEGVGNIANFELDQPYLPQOTFKYTIASKDYPEVSYDGTFF 561
QY 680 TVPTSLAYKMASOTIFYPFHAGDTYLVRVNPQFAVPKGTDALYRVFDEPHGNAYLENNYKV 739
DB 562 TVPTSLAYKMASOTIFYPFHAGDTYLVRVNPQFAVPKGTDALYRVFDEPHGNAYLENNYKV 621
QY 740 GEIKLPIPKLNOGTTTTRAGNKIPVTMANAYLDNQSTYIVEYPTILEKENOTDKPSILPOF 799
DB 622 GEIKLPIPKLNOGTTTTRAGNKIPVTMANAYLDNQSTYIVEYPTILEKENOTDKPSILPOF 681
QY 800 KRKAQENSKLDEKVEPKTSEKVEKLSSETGNSTNSNLTLEEVPTVDPVQEKVAKFAES 859
DB 682 KRKAQENSKLDEKVEPKTSEKVEKLSSETGNSTNSNLTLEEVPTVDPVQEKVAKFAES 741
QY 860 YGMKLENVLFNMDGTTIELYLPSEGVKIKKNMADPTGEAPOGNGENKPSNGKYSTGTVENQ 919
DB 742 YGMKLENVLFNMDGTTIELYLPSEGVKIKKNMADPTGEAPOGNGENKPSNGKYSTGTVENQ 801
QY 920 PTENKPADSLPEAPNEKPKVPENSTONGMLNPGCNVGSOPMLDPALEEAAPVDPVQEKLE 979
DB 802 PTENKPADSLPEAPNEKPKVPENSTONGMLNPGCNVGSOPMLDPALEEAAPVDPVQEKLE 861
QY 980 KFTASYGLGLDSVIFNMDGTTIELRLPSGEVKKNLSDFIA 1019
DB 862 KFTASYGLGLDSVIFNMDGTTIELRLPSGEVKKNLSDFIA 901

```

Search completed: May 13, 2003, 13:59:11
Job time : 35.1868 secs